

 | UNIVERSIDAD
SEÑOR DE SIPÁN

**FACULTAD DE INGENIERÍA, ARQUITECTURA Y
URBANISMO**

**ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA
DE SISTEMAS**

TESIS

**COMPARACIÓN DE TÉCNICAS DE ESTIMACIÓN
DEL GRADO DE SALINIDAD EN SUELOS CON
CONTENIDO DE HUMEDAD MEDIANTE EL
PROCESAMIENTO DE IMÁGENES
MULTIESPECTRALES**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO
DE SISTEMAS**

Autor:

Bach. Amaya Díaz Christian Omar

Asesor:

Ing. Mejía Cabrera Heber Iván

**Línea de Investigación:
Ciencias de la computación**

**Pimentel - Perú
Año 2019**

DEDICATORIAS

A mis padres.

Que sin ellos no hubiera logrado mi meta de ser Profesional.
Mamá, gracias por estar a mi lado en esta etapa de mi titulación, tu apoyo moral y entusiasmo que me brindaste para seguir adelante en mis propósitos. **Papá (q.e.p.d.)**, por el tiempo que estuviste conmigo, compartiendo tus experiencias, conocimientos y consejos, por tu amor, Gracias.

A mis maestros.

Por el tiempo y esfuerzo que dedicaron a compartir sus conocimientos, sin su instrucción profesional no habría llegado a este nivel. Quienes brindaron dedicación al impartir su cátedra de tal forma que lo aprendido sea utilizado en la vida real, por el apoyo brindado, Gracias.

A Dios.

Por darme vida, salud y sabiduría a lo largo del estudio en Ingeniería de Sistemas.

AGRADECIMIENTO

Expreso mi agradecimiento a mi asesor Ing. Heber Iván Mejía Cabrera que me brindo todo su apoyo para poder realizar esta tesis.

Agradezco a mi primo Ing. Jimmy Diaz Vega por brindarme su apoyo todo este tiempo, ya que me enseñó lo que era Matemáticas, y me encamino a seguir adelante.

Agradezco a mis padres Jorge Amaya Rojas y Marina Diaz Loaiza, y a mis hermanos, por brindarme su amor y apoyo en esta etapa de mi vida. Ya que me enseñaron a nunca rendirse, ni dejar las cosas a medias.

Gracias a todos por sus halagos.

TABLA DE CONTENIDO

I. INTRODUCCION.....	10
1.1. Realidad Problemática	10
1.2. Antecedentes de Estudio.	13
1.2.1. Antecedentes de la investigación.....	13
1.2.2. Estado de arte.	14
1.3. Teorías relacionadas al tema	20
1.3.1. Conductividad eléctrica (CE) y salinidad	20
1.3.2. Radiación electromagnética	24
1.3.3. Espectro electromagnético	25
1.3.4. Radiancia.....	27
1.3.5. Radiancia directa	27
1.3.6. Radiancia proveniente del entorno.....	28
1.3.7. Reflectancia	29
1.3.8. Teledetección.....	30
1.3.9. Métricas de medición de técnicas.	31
1.3.10. Técnicas de estimación	33
1.4. Formulación del problema.....	36
1.5. Justificación e importancia del problema	36
1.6. Hipótesis	37
1.7. Objetivos	37
1.7.1. Objetivo General	37
1.7.2. Objetivo Específicos	37
II. MATERIAL Y METODOS.....	38
2.1. Tipo y diseño de investigación	38
2.1.1. Tipo de investigación.....	38
2.1.2. Diseño de investigación.....	38
2.2. Población y muestra	38
2.3. Variables	38
2.3.1. Variable Independiente.....	38
2.3.2. Variable Dependiente	38
2.3.3. Operacionalización	39
2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.....	40



2.4.1.	Abordaje Metodológico.....	40
2.4.2.	Técnicas de Recolección de datos.....	40
2.5.	Procedimientos de análisis de datos.....	40
2.5.1.	Análisis estadístico de datos.....	40
2.6.	Criterios éticos.....	41
2.7.	Principios de rigor científico.....	41
2.8.	Método Propuesto.....	42
2.8.1.	Seleccionar las técnicas de estimación a utilizar.....	43
2.8.2.	Adquisición de imágenes multiespectrales del área de estudio.....	43
2.8.3.	Realizar calibraciones atmosféricas a las imágenes multiespectrales adquiridas. 46	
2.8.4.	Seleccionar indicadores de salinidad.....	49
2.8.5.	Recolectar muestras de suelos con contenido de humedad y realizar la prueba de conductividad eléctrica.....	52
2.8.6.	Evaluar las técnicas de estimación.....	56
III.	RESULTADOS.....	62
3.1.	Resultados en tablas y figuras.....	62
IV.	CONCLUSIONES Y RECOMENDACIONES.....	64
4.1.	Conclusiones.....	64
4.2.	Recomendaciones.....	65
	REFERENCIAS.....	66
	ANEXOS.....	68



INDICE DE FIGURAS

Figura 1 : Clasificación del espectro electromagnético	25
Figura 2: Espectro visible por el ojo humano	26
Figura 3: Componente De La Radiancia	29
Figura 4: Proceso de calibración radiométrica y geométrica de imágenes	31
Figura 5: Regresión Lineal Múltiple	34
Figura 6: Random Forest	35
Figura 7: Método Propuesto de la tesis “Comparación De Técnicas De Estimación Del Grado De Salinidad En Suelos Con Contenido De Humedad Mediante	42
Figura 8: Imágenes Multiespectrales descargadas de Landsat 8	45
Figura 9: Información de las Imágenes Multiespectrales descargadas de Landsat 8.....	45
Figura 10: Imágenes De Reflectancia Generadas Al Procesar Las Imágenes	48
Figura 11: Imágenes Procesadas Con Los Indicadores Antes Mencionados	52
Figura 12: Resultados De La Estimación Usando Regresión Lineal Múltiple	60
Figura 13: Resultados De La Estimación Usando Random Forest	61
Figura 14: Resultado Del (R² Y RMSE) del Random Forest	62
Figura 15: Resultado Del (R² Y RMSE) de la Regresión Lineal Múltiple	62
Figura 16: Resultado Del R² - Comparando Random Forest Y Regresión Lineal Múltiple	63
Figura 17: Resultado Del RMSE - Comparando Random Forest Y Regresión Lineal Múltiple	63
Figura 18: Búsqueda De Imágenes Multiespectrales Por Lugar.....	69
Figura 19: Búsqueda De Imágenes Multiespectrales Por Fecha.	70
Figura 20: Búsqueda De Imágenes Multiespectrales Por Satélite.	70
Figura 21: Búsqueda De Imágenes Multiespectrales Por Filtro De Porcentaje De Nubosidad.	71
Figura 22: Vista De Adquisición De Imágenes Multiespectrales.	71
Figura 23: Adquisición De Paquete De Imágenes Multiespectrales.	72
Figura 24: Archivo De Metadata De Un Paquete De Imágenes Multiespectrales.....	72
Figura 25: Variables De Metadata A Utilizar Para Calcular La Reflectancia	73
Figura 26: Código Fuente En Python Para Leer El Archivo De Metadata	74
Figura 27: Código Fuente En Python De La Función Para Leer Una Imagen	75
Figura 28: Código Fuente En Python Para Calcular La Reflectancia.....	75
Figura 29: Código Fuente En Python Para Calcular La Reflectancia.....	75
Figura 30: Código Fuente En Python De La Ejecución De Las Funciones De Leer De Imagen Y Leer La Metadata	76
Figura 31: Código Fuente En Python De Lectura De Una Imagen Espectral Procesada	77
Figura 32: Código fuente en Python implementación del indicador NDVI	77
Figura 33: Código Fuente En Python De Lectura De Los Puntos De Origen Y El Tamaño De Pixel De Una Imagen Espectral	77
Figura 34: Código Fuente En Python De La Exportación De Un Indicador Calculado.....	78
Figura 35: Código Fuente En Python De Lectura De Un Indicador.....	80
Figura 36: Código Fuente En Python De Lectura De Coordenada, Conversión Y Búsqueda De Coordenada.....	80
Figura 37: Código Fuente En Python De Ejecución De Búsqueda De Coordenadas Y Extracción De Indicadores.....	81
Figura 38: Código Fuente En Python De Ejecución De Búsqueda De Coordenadas Y Extracción De Indicadores.....	81



Figura 39: Código Fuente En Python De Ejecución De La Técnica Regresión Lineal Múltiple82
 Figura 40: Código Fuente En Python De Ejecución De La Técnica Random Forest82
 Figura 41: Análisis De Muestras De Suelos Con Humedad, Mediante Conductividad Eléctrica.....83

INDICE DE TABLAS

Tabla 1	<i>Superficie dañada con problemas de drenaje y salinidad en la Costa Peruana</i>	11
Tabla 2	<i>Unidades de la Conductividad Eléctrica</i>	11
Tabla 3	<i>Clasificación de salinidad de suelos, mediante a la respuesta de los cultivos.</i>	12
Tabla 4	<i>Clasificación de la salinidad.</i>	21
Tabla 5	<i>Cultivos hortícolas</i>	22
Tabla 6	<i>Cultivos leñosos</i>	23
Tabla 7	<i>Alcance de los suelos afectados por la sal</i>	24
Tabla 8	<i>Límites del espectro electromagnético según su división</i>	26
Tabla 9	<i>Límite de la luz visible según los colores RGB</i>	27
Tabla 10	<i>Resolución espacial de satélites</i>	30
Tabla 11	<i>Bandas del sensor OLI del satélite Landsat 8</i>	44
Tabla 12	<i>Descripción de los archivos Landsat 8</i>	46
Tabla 13	<i>Imágenes espectrales de reflectancia, banda y su identificador</i>	48
Tabla 14	<i>Correlación entre los indicadores de salinidad, vegetación y bandas, y el valor de conductividad eléctrica</i>	49
Tabla 15	<i>Correlación entre los indicadores de salinidad, vegetación, y el valor de conductividad eléctrica</i>	50
Tabla 16	<i>Ubicación de las muestras tomadas y su respectivo valor de conductividad eléctrica</i>	53
Tabla 17	<i>Relación entre los indicadores elegidos y el valor de conductividad eléctrica</i>	55
Tabla 18	<i>Coefficientes</i>	56
Tabla 19	<i>Resultado De La (Y) Estimada y (Y) Real</i>	57
Tabla 20	<i>Resultado Del RMSE</i>	58
Tabla 21	<i>Promedio del Y (Real)</i>	58
Tabla 22	<i>Resultado Del R2</i>	59
Tabla 23	<i>RESULTADOS DE LA ESTIMACION USANDO REGRESIÓN LINEAL MÚLTIPLE</i>	60
Tabla 24	<i>RESULTADOS DE LA ESTIMACION USANDO RANDOM FOREST</i>	61
Tabla 25	<i>Población de técnicas de estimación</i>	68
Tabla 26	<i>Información de coordenada a buscar y el indicador donde se buscará</i>	79
Tabla 27	<i>Ubicación y valor de un pixel en una imagen espectral</i>	79



RESUMEN

El fenómeno de la salinidad es un problema clásico de degradación de la tierra que enfrenta el hombre, al dar un uso deficiente a los fertilizantes y a la calidad de riego, esto puede ocasionar peligro en la agricultura, por lo que tenemos que controlar y prevenir los suelos salinos, los suelos que están afectados por la salinidad tienen problemas en el crecimiento y desarrollo de cultivos, sin embargo, hay suelos que presentan salinidad que a su vez generan productividad ya que estos son diagnosticados y mejorados sobre la magnitud del problema (Otero et al., 2008). El método tradicional para estimar la salinidad de los suelos es la conductividad eléctrica (CE), pero estos son muy costosos y requieren mucho tiempo, la CE se toma a una temperatura de 18°C o 25°C (SEMARNAT, 2010), este proyecto de investigación presenta un estudio de comparación entre técnicas de estimación Regresión Lineal Múltiple (RLM) y Random Forest (RF) para la estimación de salinidad en suelos con contenido de humedad, los valores obtenidos mediante indicadores de salinidad de imágenes multiespectrales se utilizaron para estimar los valores de CE. El resultado que se muestra en la estimación de salinidad en suelos con contenido de humedad se obtuvo que el RLM es más preciso que el RF con un coeficiente de determinación (R^2) de 0.86 y 0.73; y un error cuadrático medio (RMSE) de 0.11 dS/m y 0.15 dS/m respectivamente.

Palabras Claves: Salinidad en suelos, conductividad eléctrica, Imágenes Multiespectrales, Regresión Lineal Múltiple, Random Forest

ABSTRACT

The phenomenon of salinity is a classic problem of land degradation that man faces, by giving poor use of fertilizers and irrigation quality. This can cause danger in agriculture, so we have to control and prevent saline soils, soils affected by salinity have problems in the growth and development of crops. However, there are soils that have salinity that in turn generate productivity, since they are diagnosed and improved in the magnitude of the problem (Otero et al., 2008). The traditional method to estimate soil salinity is electrical conductivity (EC), but these are very expensive and time-consuming. The EC is taken at a temperature of 18°C or 25°C (SEMARNAT, 2010). This research project presents a comparative study between estimation techniques of Multiple Linear Regression (RLM) and Random Forest (RF) for the estimation of salinity in soils with moisture content. The values obtained through salinity indicators of multispectral images were used to estimate the EC values. The result shown in the salinity estimation in soils with moisture content was that the RLM is more accurate than the RF with a coefficient of determination (R^2) of 0.86 and 0.73; and a mean square error (RMSE) of 0.11 dS/m and 0.15 dS/m respectively.

Keywords: soil salinity, electrical conductivity, multispectral images, multiple linear regression, random forest



I. INTRODUCCIÓN

1.1. Realidad Problemática

El fenómeno de la salinidad es uno de los problemas clásicos de degradación de la tierra que ha tenido que enfrentar el hombre, al darle mal uso a los fertilizantes y a la calidad de riego, esto puede causar peligro en la agricultura, es por ello que se tiene que controlar y prevenir los suelos afectados por la salinidad, así mismo los suelos que están afectados por la salinidad es posible que tenga problemas en el crecimiento y desarrollo de cultivos, sin embargo hay suelos que presentan salinidad que a su vez generan productividad ya que estos son diagnosticados y mejorados sobre la magnitud del problema (Otero et al., 2008).

Por otro lado, en un clima seco, el agua freática emerge por capilares a la superficie y se evapora, y las sales se acumulan. Cuando dejamos de regar un campo por varios años seguidos, el suelo permanece con bastante salinidad y no apto para el uso agrícola (ALVA et al., 1976).

El problema de la salinidad en nuestra costa peruana, es causado por el nivel freático elevado, poco disponible en agua de riego y baja calidad del agua. En el caso que el agricultor riegue sus cultivos hasta humedecer el suelo, y al ser consumidas por las plantas, se almacenan las sales y afectan los cultivos en diferentes grados, y el suelo se mantendrá a un nivel alto de salinidad ya sea por la mala calidad del agua o aun si el agua es salina y escasa (ALVA et al., 1976).

Se realizó una evaluación de drenaje y salinidad a 42 de los 52 valles de nuestra costa peruana obteniendo en total un terreno de 757,000 ha; las cuales el 501,780 ha (66.5%) no se encontraron problemas de drenaje y salinidad, y el 255,230 ha (34%) si presenta diversos grados de drenaje y salinidad (ALVA et al., 1976).



Tabla 1

Superficie dañada con problemas de drenaje y salinidad en la Costa Peruana

Grado del problema	Extensión (ha)	%
Sin problemas.	501,780	66.5
Ligero problemas de drenaje y/o salinidad.	102,360	13.5
Moderados a fuertes problemas de salinidad.	19,385	2.5
Moderados a fuertes problemas de salinidad y drenaje.	133,485	17.5

Nota: ha = Hectáreas, % = Porcentaje

Fuente: Problemas de drenaje y salinidad en la Costa Peruana.

Mediante un estudio realizado la salinización de los suelos es un proceso de degradación ecológica donde el suelo pierde total o parcialmente la producción, esto afecta en todo el mundo. A raíz de todo, esto vendría ser un peligro ambiental que amenaza la vida de los ecosistemas. Para lograr una mejor producción es necesario desarrollar una serie de estudios para estimar el nivel del sal y contenido de humedad en los suelos (Yang & Yu, 2017), debido que cada cultivo tiene cierto grado de sensibilidad a la sal, que permitan su desarrollo, se sugiere realizar monitoreo y evaluación de salinidad en los suelos para luego gestionar y prevenir el aumento de sal en los suelos (Ding & Yu, 2014).

Es por ello, el método tradicional para estimar la salinidad de los suelos es realizada con la conductividad eléctrica, pero estos son muy costosos y requieren mucho tiempo, la conductividad eléctrica se toma a una temperatura de 18°C o 25°C (SEMARNAT, 2010). Para realizar la medida de la conductividad eléctrica (CE) se tiene que realizar sobre el agua o extraer una parte del suelo humectado. Las unidades que utiliza la CE es el simen/metro o subunidades como el decisiemen/metro (ds/m) (Badia Villas, 1992).

Tabla 2

Unidades de la Conductividad Eléctrica

1 ds/m	Es igual a	1 mmho/cm
1 mmho/cm	Es igual a	1000 umho/cm

Nota: ds = decisiemen, mmho = milimho, umho = micromhos, m = metros, cm = centímetros

Fuente: Suelos Afectados por sales



Clasificamos la salinidad de los suelos en 5 clases, siguiendo la conductividad eléctrica del extracto acuoso (Gobierno del Estado de Baja California, 2009).

Tabla 3

Clasificación de salinidad de suelos, mediante a la respuesta de los cultivos.

Tipo de Suelo	Salinidad	C.E. (ds/m)	Efectos
Normales	Muy ligera	0-2	Casi nulos
	Ligera	2-4	Puede afectar a los cultivos sensibles
Salinos	Media	4-8	La mayoría de los cultivos son afectados
	Fuerte	8-16	Pueden prosperar cultivos tolerantes
	Muy Fuerte	>16	Se desarrollan cultivos muy tolerantes

Nota: ds = decisiemen, m = metros, C.E. = Conductividad Eléctrica

Fuente: Estudio bibliográfico sobre la salinidad del valle de Mexicali

Para poder detectar la salinidad en los suelos con contenido de humedad utilizaremos procesamiento de imágenes multiespectrales. Las imágenes multiespectrales, proporcionan una amplia información espectral para identificar y distinguir materiales espectralmente únicos. El análisis de imagen se ha convertido en las tecnologías más potentes y de ágil crecimiento en el campo de teledetección. Es una técnica avanzada para obtener un gran número de bandas de longitud de onda, medidas en un solo píxel (Nila Rekha, Gangadharan, Pillai, Ramanathan, & Panigrahi, 2012).

Con el desarrollo y la perfección de las tecnologías de teledetección hiperespectral, ha sido la principal técnica aplicada para el mapeo de minerales (Clark, 1999) para detectar las propiedades del suelo, incluida la humedad, el contenido orgánico y la salinidad (Bendor, 2000).



Se sabe que el problema del mapeo de suelos afectados por sal es una cuestión difícil ya que es un proceso dinámico. La salinización del suelo es un proceso en el que las sales disueltas se agrupan en la superficie del suelo y en las capas superiores del suelo. Sin embargo, el uso del mapeo de análisis hiperespectral y el monitoreo de tierras degradadas, especialmente en suelos afectados por sal, han mostrado una gran promesa de mayor velocidad, precisión y rentabilidad (Mermut y Eswaran, 2001).

Es por ello que, para solucionar el problema de la detección de la salinidad en suelos con contenido de humedad, tenemos que comparar las técnicas de **Random Forest** y **Regresión Lineal Múltiple**, que estimaran el grado de salinidad mediante procesamiento imágenes multiespectrales con las muestras de suelo que se harán mediante conductividad eléctrica en el laboratorio, que servirá para que los agricultores puedan sembrar sin ser afectados por la salinidad.

1.2. Antecedentes de Estudio.

1.2.1. Antecedentes de la investigación.

(Fox Llenera, 2013) desarrollo el trabajo de investigación *“Evaluación de pérdida de suelo por salinización en la parte baja de la cuenca del Jequetepeque: san pedro de Lloc (1980 – 2003)”* en la Universidad Católica del Perú, el problema que enfrento fue que la salinización de los suelos reduce la productividad agrícola, propuso como solución diagnosticar los suelos afectados por la salinización en San Pedro de Lloc. Desarrolló la solución caracterizando los deterioros de suelos agrícola, inspeccionar las propiedades de los suelos con una evaluación de parámetros físicos y químicos, mapear el tamaño de las zonas agrícolas, realizar una base de datos georreferenciados que ayuden al monitoreo del deterioro de los suelos, de las 23 muestras que se obtuvo, la muestra número 2s presenta un mayor grado de conductividad eléctrica de 2.63 ds/m. Se concluyó que el terreno del subsector bajo, contiene gran cantidad de salinidad, a diferencia de los demás subsectores que se localizan en el área de estudio.



1.2.2. Estado de arte.

(Palomo et al., 2013) desarrollaron el trabajo de investigación *“Uso de aguas salinas y frecuencias de aplicación asociado a la salinización de suelos y producción de algodónero”* en el Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) de México, el problema que enfrento es la capacidad de producción y estabilidad de la planta de algodón, es afectada por las condiciones de salinidad ya sea del suelo o el agua de riego. Se propuso realizar pruebas para observar que efectos causa las diferentes condiciones de salinidad del agua y riego a las plantas de algodón. Estas pruebas se realizaron en muestras con 4 niveles distintos de salinidad del agua de riego (1.12, 3.08, 6.43 y 10.27 dS m), en distintas frecuencias de riego (2, 3, 4 y 5) respectivamente, todas las muestras fueron puestas en las mismas condiciones de suelo (0.60 m de diámetro y 0.40 m de profundidad). Concluyendo las pruebas se observó que la producción de las plantas de algodón bajo las condiciones de salinidad de agua (1.12, 3.08 y 6.43 dS m) no se vieron afectadas, sin embargo, las plantas bajo las condiciones de salinidad en el agua de 10.27 dS m se vieron afectadas en su producción con un decaimiento en un 23% en comparación a las otras muestras, determinando que la plantación de algodón no es afectada cuando está expuesta a una salinidad menor a 7.7 dS m, ya que este es el límite de tolerancia de las plantas de algodón.

(Dasgupta et al., 2015) desarrollaron el trabajo de investigación *“El cambio climático y la salinidad del suelo: el caso de la costa de Bangladesh”* en la Real Academia de las Ciencias de Suecia, el presente proyecto se realiza con la finalidad de estimar el nivel de salinidad en los suelos costeros de Bangladesh en el año 2050, para ello se iniciará por evaluar información registrada en el periodo 2001 – 2009 de estaciones de monitoreo del suelo, a través de estos datos se estimaría un modelo econométrico espacial. Al estudiar los cambios ocasionados por el clima en la salinidad de los ríos y las proyecciones de precipitaciones en base a las tendencias temporales, posteriormente se obtuvo los datos sobre los cambios de salinidad registrados en el periodo 2001 – 2009, a partir de esto se estimaría un modelo econométrico espacial que une la salinidad del suelo, la salinidad de los ríos cercanos, elevación de la tierra, temperatura y las precipitaciones. Mediante los cambios climáticos inducidos por el clima, salinidad de los ríos, precipitaciones y la temperatura, se



proyectará las estimaciones. Según los resultados de las 41 estaciones de control, se llegó a estimar que en promedio se estima que el cambio de la salinidad del suelo para el año 2050 es de 39%, sobre la mediana el 25% de las estaciones estimaron que la salinidad del suelo sería de 51% o superior.

(**Kayri, Kayri, & Tunay Gencoglu, 2017**) desarrollaron el trabajo de investigación *“The Performance Comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using Photovoltaic and Atmospheric Data”*, en el departamento de Electrical and Electronic Engineering, el proyecto tiene como finalidad que los resultados de la estimación por Regresión Lineal Múltiple, Bosques Aleatorios y Red Neuronal Artificial se examinan comparativamente, para la comparación de estas técnicas de extracción de datos, los datos de producción de energía de un módulo fotovoltaico fueron usados en la investigación. En el estudio, el modelo se constituyó en siete variables. Una de las variables es dependiente (potencia) y las otras son variables independientes (radiación global, temperatura, velocidad del viento, dirección del viento, humedad relativa, ángulo de elevación solar). Para ello se utilizó el error absoluto promedio y el coeficiente de correlación para comparar el rendimiento de estimación de las técnicas de extracción de datos mencionadas. Donde el coeficiente de correlación es 0.963 en la técnica de Regresión Lineal Múltiple, el coeficiente de correlación es 0.986 en la técnica de Bosques Aleatorios, y el mayor coeficiente de correlación se obtuvo en la estructura de la Red Neuronal Artificial ($R^2 = 0.997$). De acuerdo con los tres métodos de extracción de datos, la radiación global se encontró como el predictor más importante. Si bien el predictor menos importante es la dirección del viento tanto en la técnica de Red Neuronal Artificial como en la técnica de Bosques Aleatorios, el ángulo de elevación solar es el predictor menos importante en la técnica de Regresión Lineal Múltiple.

(**Mousavi, Habibnejad, Kavian, Solaimani, & Khormali, 2017**) desarrollaron el trabajo de investigación *“Digital Mapping of Topsoil Salinity Using Remote Sensing Indices in Agh-Ghala Plain, Iran”* en Sari University of Agricultural Science and Natural Resources, Sari de Iran, el proyecto presenta que la salinización del suelo es un proceso mundial de degradación de la tierra en regiones áridas y semiáridas que conlleva graves



consecuencias económicas y sociales, se realizó los análisis de salinidad del suelo mediante dos modelos estadísticos lineales (Regresión Lineal Múltiple) y no lineales (Red Neuronal Artificial) utilizando datos de Landsat OLI en la llanura de Agh-Ghala ubicada en el noreste de Irán, también se resolvió la conductividad eléctrica (CE) in situ del suelo de 156 muestras de suelo (profundidad de 0-15 cm), Se usó una correlación de Pearson entre 26 índices espectrales derivados de datos OLI de Landsat y ECs medidos in situ para aplicar índices eficientes en la evaluación de la salinidad del suelo. Los índices mejor correlacionados, tales como bandas azules, verdes y rojas, índices de intensidad (Int1, Int2), índices de salinidad del suelo (Si1, Si2, Si3, Si11, Aster-Si), índices de vegetación (NDVI, DVI, RVI, SAVI), Se utilizaron índices de verdor y humedad para desarrollar dos modelos. La comparación entre dos modelos de estimación mostró que el rendimiento del modelo ANN ($R^2 = 0.964$ y $RMSE = 2.237$) fue más confiable que el del modelo MLR ($R^2 = 0.506$ y $RMSE = 9.674$) en el monitoreo y la predicción de la salinidad del suelo, del área total, el 66% y el 55.8% se identificaron como no salinos, ligeramente y muy ligeramente salinos para los modelos ANN y MLR, respectivamente. Esto muestra que los datos de la detección remota se pueden utilizar de manera efectiva para modelar y mapear variaciones espaciales de salinidad del suelo

(Allbed, Kumar, & Sinha, 2014) en su artículo científico titulado *“Mapping and Modelling Spatial Variation in Soil Salinity in the Al Hassa Oasis Based on Remote Sensing Indicators and Regression Techniques”*, realizado en el Department of Ecosystem Management, School of Environmental and Rural Science en la Universidad de New England. Encontraron, que la salinidad del suelo causa problemas ambientales más dañinos del mundo, especialmente en las regiones áridas y semiáridas. Un enfoque integrado que utiliza sensores remotos, además de varios métodos estadísticos, ha demostrado tener éxito desarrollando técnicas de predicción de salinidad del suelo, la finalidad de este estudio es desarrollar técnicas de regresión estadística basados en indicadores de detección remota para predecir y mapear la variación espacial de la salinidad del suelo en el oasis de Al Hassa. Se calcularon diferentes índices espectrales a partir de bandas originales de imágenes de IKONOS. La correlación estadística entre las mediciones de campo de la conductividad eléctrica (EC), los índices espectrales y las bandas originales de IKONOS mostraron que el



Índice de salinidad (SI) y la banda roja (banda 3) tenían la mayor correlación con la CE. La combinación de estas dos variables detectadas de forma remota en un solo modelo produjo el mejor ajuste con $R^2 = 0,65$. Los resultados revelaron que el alto rendimiento de este modelo combinado se atribuye a: (i) la resolución espacial de las imágenes; (ii) el gran potencial de las imágenes mejoradas, derivadas de SI, al realzar y delinear la variación espacial de la salinidad del suelo; y (iii) la superioridad de la banda 3 en la recuperación de características y patrones de salinidad del suelo, lo que se explica por la alta reflectancia de la corteza de la superficie lisa y brillante y la baja reflectancia de la corteza gruesa y oscura. Los mapas de salinidad del suelo generados utilizando el modelo seleccionado mostraron que los suelos fuertemente salinos ($>16\text{dS/m}$) con distribución espacial variable fueron la clase dominante en el área de estudio. La variabilidad espacial de esta clase sobre las áreas investigadas se atribuyó a una variedad de factores, incluidos los factores del suelo, los factores relacionados con la gestión y los factores climáticos. Los resultados demuestran que el modelado y el mapeo de la variación espacial en la salinidad del suelo basado en el análisis de regresión y los datos de detección remota es un enfoque prometedor, ya que facilita la detección oportuna con un procedimiento de bajo costo y permite a los encargados de tomar decisiones decidir qué medidas deben tomarse en el proceso. etapas tempranas para evitar que la salinidad del suelo se vuelva prevalente, manteniendo las tierras agrícolas y los ecosistemas naturales.

(Shahabi, Jafarzadeh, Neyshabouri, Ghorbani, & Kamran, 2017) en su investigación titulado “*Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods*”. Encontraron, que la salinidad como una propiedad importante del suelo que representa la reducción de la fertilidad en el mundo. La información precisa sobre el cambio espacial de la salinidad del suelo es esencial para el manejo y la utilización sostenibles del suelo en tierras agrícolas. Para ello se tomaron 150 muestras de suelo de Dashte-e-Tabriz Irán y se analizaron, y la salinidad del suelo se estimó mediante parámetros de superficie terrestre, incluidos la elevación, el aspecto, la longitud de la pendiente, el índice de humedad, la pendiente y el índice de diferencia de vegetación normalizada como parámetros básicos. Para modelar y predecir la salinidad, se utilizaron el Krigeaje Ordinario (OK), Redes Neuronales Artificiales (ANN) y Regresiones



Lineales Múltiples (MLR). La precisión de los modelos se evaluó mediante el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE). Sobre la base de la correlación de Pearson, la elevación, la diferencia de normalización y los índices de humedad se seleccionaron para la modelación espacial de la salinidad del suelo a partir de seis parámetros de la superficie terrestre. Los resultados mostraron que la ANN tenía el RMSE más bajo y el R^2 más alto. Los valores de R^2 , RMSE y MAE fueron 0.36, 25.89 y 17.06 para la regresión y 0.56, 17.70 y 13.05 para KO y 0.69, 16.06 y 11.60 para ANN, respectivamente, lo que indica una mayor precisión de ANN en comparación con MLR y KO.

(Yue, Feng, Yang, & Li, 2018) en su trabajo de investigación titulado “*A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy*”. Encontraron que, la biomasa sobre el suelo (AGB) proporciona un vínculo vital entre el consumo de energía solar y el rendimiento, por lo que su estimación correcta es crucial para controlar con precisión el crecimiento de los cultivos y predecir el rendimiento. Para estimar AGB se utilizó 54 índices de vegetación (por ejemplo, Índice de Vegetación de Diferencia Normalizada, Índice de Vegetación Ajustada al Suelo) y ocho técnicas de regresión estadística: Red Neuronal Artificial (ANN), regresión lineal multivariable (MLR), regresión de árbol de decisión (DT), árbol de regresión binaria reforzada (BBRT), regresión de mínimos cuadrados parciales (PLSR), regresión de Bosques Aleatorios (RF), regresión de máquina de vectores de soporte (SVM) y regresión de componente principal (PCR), que se utilizan para analizar datos hiperespectrales adquirido mediante el uso de un espectrofotómetro de campo. Los índices de vegetación (VI) determinados a partir de los espectros se utilizaron primero para entrenar técnicas de regresión para el modelado y la validación para seleccionar la mejor entrada de VI, y luego se sumaron con ruido gaussiano blanco para estudiar cómo los errores de detección remota afectan las técnicas de regresión. Los VI se dividieron en grupos de diferentes tamaños utilizando varios métodos de muestreo para modelar y validar para probar la estabilidad de las técnicas. Finalmente, el AGB se estimó utilizando una validación cruzada de dejar-uno-fuera con estas poderosas técnicas. Los resultados del estudio demuestran que, de las ocho técnicas investigadas, PLSR y MLR se desempeñan mejor en términos de estabilidad y son



más adecuadas cuando se requieren estimaciones de alta precisión y estables de relativamente pocas muestras. Además, la RF es extremadamente robusta contra el ruido y es más adecuada para lidiar con observaciones repetidas que involucran datos de sensores remotos (es decir, datos afectados por la atmósfera, las nubes, los tiempos de observación y / o el ruido del sensor). Finalmente, el método de validación cruzada de dejar fuera uno indica que PLSR proporciona la mayor precisión ($R^2 = 0,89$, $RMSE = 1,20$ t / ha, $MAE = 0,90$ t / ha, $NRMSE = 0,07$, $CV (RMSE) = 0,18$); por lo tanto, PLSR es más adecuado para trabajos que requieren modelos de estimación de alta precisión. Los resultados indican que todas estas técnicas proporcionan una precisión impresionante. La comparación y el análisis aquí proporcionados revelan las ventajas y desventajas de las técnicas ANN, MLR, DT, BBRT, PLSR, RF, SVM y PCR y pueden ayudar a los investigadores a construir modelos de estimación de AGB eficientes.

(Yang & Yu, 2017) en su trabajo de investigación titulado *“Estimación de la salinidad del suelo bajo diversas condiciones de humedad: un estudio experimental”*. El problema que encontraron fue que la salinización del suelo es un peligro ambiental ciertamente grave que provoca fenómenos de degradación del suelo, como el sellado del suelo, cambios estructurales y formación de costras, la reducción del rendimiento de los cultivos y la producción agrícola, y amenaza la estabilidad de los ecosistemas. Además, hubo un área de 77-Mha salinizada como consecuencia de las actividades humanas y las áreas afectadas por la sal, las cuales seguirán aumentando debido al aumento de la presión demográfica. Propusieron, tener un método rápido y rentable para identificar y mapear el suelo de salinidad en una gran escala de campo. Por ello, mediante tecnología de observación utilizada para identificar y mapear los efectos de sal en las tierras desarrollaron un modelo semianalítico que utiliza una función de exponente para estimar el contenido de sal del suelo (SSC) bajo diferentes niveles de humedad, basados en un experimento de laboratorio de control. Se obtuvieron como resultados, que el error de raíz cuadrada y el error relativo medio fueron 0,799 y 31,294%, respectivamente, cuando se aplicó el modelo para estimar los SSC por la reflectancia del suelo húmedo. Considerando ambos efectos de la humedad y la sal del suelo sobre la reflectancia del suelo, el modelo semianalítico redujo el error estimado de SSC. Por lo tanto, este enfoque proporciona una nueva forma de estimar la salinidad del suelo a partir de los espectros del suelo bajo diversas condiciones de humedad del suelo, y será una aplicación potencial para el mapeo de SSC a gran escala.



1.3. Teorías relacionadas al tema

1.3.1. Conductividad eléctrica (CE) y salinidad

La CE es el volumen del agua para transportar corriente eléctrica por medio de los iones diluidos y la salinidad es una medida de la abundancia de sales diluidas en agua.

La CE y la salinidad tienen una relación por que la abundancia de iones diluidos incrementa los valores de ambos.

Cuanto mayor es la abundancia de iones diluidos en agua, mayor es la conductividad de la solución del suelo. Para medir la CE depende de la temperatura de la solución y esta puede ser de referencia a los 18°C o 25°C. La CE puede ser medida en una pasta de suelo saturado o en una suspensión más diluida. (SEMARNAT, 2010)

Lo que define la calidad y fertilidad de cualquier suelo agrícola es el volumen de sales concurrentes. La salinidad es más reflejada en suelos áridos y semiáridos. La forma en que se calcula la salinidad en suelos es por medio de la CE, la CE calcula la densidad de sales solubles concurrente en la solución del suelo. Las unidades que se aplican para medir la CE son Ds/m (decisiemens por metro) que son equivalentes a las que se aplicaban en mmhos/cm. (Havlin,2013).

La conductividad eléctrica mostrada en siguiente ecuación:

$$CE_{25} = CE_t \times ft$$

Donde:

CE_t = La conductividad eléctrica medida del extracto de tierra en t °C (grados centígrados)

ft = es el valor corregido de la conductividad eléctrica en t °C (grados centígrados)



Tabla 4

Clasificación de la salinidad.

CE (dS/m)	clasificación	efecto sobre las plantas
<0.75	Liberado de sales	No hay restricción para ningún cultivo
0.75-1.15	Muy menudo en sales	No hay restricción para ningún cultivo
1.10-2.0	Muy levemente salino	Afecta el rendimiento a cultivos muy sensibles
2.0-4.0	Levemente salino	Afecta la productividad a cultivos sensibles
4.0-8.0	Medio salino	Afecta la productividad de casi todos los cultivos
8.0-12.0	Intensamente salino	Pueden desarrollarse cultivos tolerantes a la salinidad
12.0-16.0	Muy Intensamente salino	Pueden desarrollarse cultivos muy tolerantes a la salinidad
>16.0	Excesivamente salino	Ningún cultivo agrícola crece rentablemente

Nota: ds = decisiemen, m = metros, C.E. = Conductividad Eléctrica

Fuente: Salinidad del suelo

El contenido de sales que se realiza mediante la CE, nos permite utilizar el suelo para la agricultura o no hacerlo, dependiendo del sistema de riego (Castellanos,2000).

En la siguiente tabla se muestran la gran mayoría de cultivos hortícolas y leñosos donde se mostrarán los valores y clasificación del tipo de tolerancia (Maas y Hoffman 1977).

Estos parámetros debemos tomarlo como manual al momento de estimar perdidas de producción debido a la salinidad, para realizar una estimación más eficaz se debe realizar experimentos con diversos locales de cultivos y en que clima, suelo y zona se encuentran.



Tabla 5
Cultivos hortícolas

Cultivo	CE límite (a) (dS/m)	Tipo tolerancia
Cebolla	1.2	Sensible
Zanahoria	1	Sensible
Apio	1.8	Levemente sensible
Avena	1.7	Levemente sensible
Berenjena	1.1	Levemente sensible
Col Brócoli	2.8	Levemente sensible
Calabaza	1.2	Levemente sensible
Coliflor	1.8	Levemente sensible
Espinaca	2	Levemente sensible
Girasol	1.7	Levemente sensible
Lechuga	1.25	Levemente sensible
Maíz dulce	1.7	Levemente sensible
Maíz grano	1.7	Levemente sensible
Melón	1.7	Levemente sensible
Patata	1.7	Levemente sensible
Pepino	2	Levemente sensible
Pimiento	1.5	Levemente sensible
Rábano	1.2	Levemente sensible
Sandía	0.9	Levemente sensible
Tomate	2	Levemente sensible
Alcachofa	6.1	Levemente sensible
Trigo	6	Levemente sensible
Algodón	7.7	Tolerante
Cebada	8	Tolerante
Remolacha azucarera	7	Tolerante

Nota: C.E. = Conductividad Eléctrica

Fuente: Tolerancia de los cultivos a la salinidad



Tabla 6

Cultivos leñosos

Cultivo	CE límite (a) (dS/m)	Tipo tolerancia
Aguacate	1.6	Sensible
Almendro	1.5	Sensible
Cerezo	1.5	Sensible
Ciruelo	1.5	Sensible
Caqui	1.7	Sensible
Mandarino	1.7	Sensible
Manzano	1.5	Sensible
Limonero x Naranja amargo	1.5	Sensible
Melocotonero	1.7	Sensible
Naranja	1.7	Sensible
Níspero	1.6	Sensible
Peral	1.5	Sensible
Limonero	1.5	Levemente sensible
Viña	1.5	Levemente sensible
Granado	5	Levemente sensible
Higuera	4.2	Levemente sensible
Olivo	5	Levemente sensible
Palmera	4	Tolerante

Nota: C.E. = Conductividad Eléctrica

Fuente: Tolerancia de los cultivos a la salinidad

En la siguiente tabla mostraremos las zonas afectadas por la sal, que corresponde al 7% de la dimensión continental de la tierra, en los territorios afectados, unos 77 millones de hectáreas están salinizadas por efectos de las actividades humanas con el 58% de estos concentrados en áreas irrigadas (Szabolcs,1979)



Tabla 7

Alcance de los suelos afectados por la sal

Región	Millones de Hectáreas
México Y Centro América	2
América Del Norte	15.7
Sudeste De Asia	20
Europa	50.8
África	80.5
Sur De Asia	84.8
América Del Sur	129.2
Norte Y Centro De Asia	211.7
Australia	357.3
Total	952

Fuente: Detección de la Salinidad en los Suelos de Cultivos de Algodón usando Imágenes de Satélite en la Costa Norte del Perú

1.3.2. Radiación electromagnética

Es la combinación de los vectores de campo eléctrico y magnético que se propagan por el espacio a través de ondas portadoras de energía. En cual existen ondas electromagnéticas que se extienden por medio del espacio a la velocidad de la luz (Tipler, 2000). Esta forma formada por la ecuación:

$$c = \lambda * F$$

Donde:

λ = Longitud de onda

F = Frecuencia



Y Según la teoría cuántica, se basa en propiedades energéticas:

$$Q = h * F$$

Donde:

h = Es la constante de Planck

F = Frecuencia

1.3.3. Espectro electromagnético

Se propaga de los rayos cósmicos con longitudes de onda hasta las radiofrecuencias bajas y longitudes de onda larga. Las longitudes de ondas más usadas son de 300 y 400 nanómetros. La región visible e infrarrojo cercano entre 400 nm y um son las más empleadas. Con el que más estamos familiarizados es con el espectro visible ya que puede ser vista cuando la luz blanca es esparcida por gotitas de agua en un arco iris. (Slater,1980).

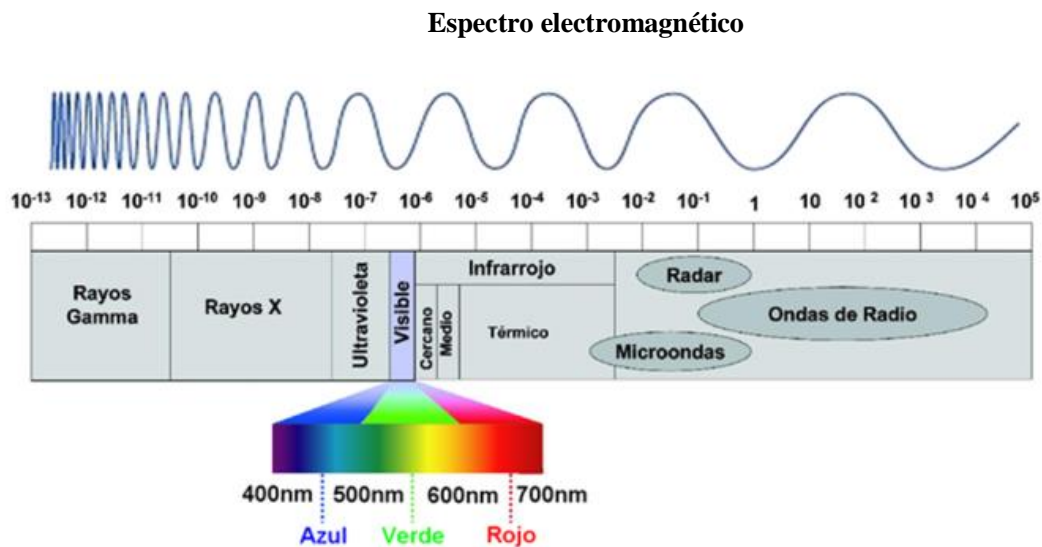


Figura 1 : Clasificación del espectro electromagnético

Fuente: Radio Frecuencia España



Tabla 8

Límites del espectro electromagnético según su división

DIVISION	LIMITES
Rayos Gamma	<0.03nm
Rayos X	0.03–300 nm
Radiación ultravioleta	0.30-0.38 μm
Luz Visible	0.38–0.72 μm
Infrarrojo cercano	0.72–1.30 μm
Infrarrojo medio	7.0–1,000 μm (1 mm)
Radiación microondas	1 mm–30 cm
Ondas de Radio	≥ 30 cm

Nota: nm = nanómetro, μm = micrómetros, cm. = centímetros

Fuente: Introduction to Remote Sensing Fift Edition

Espectro visible

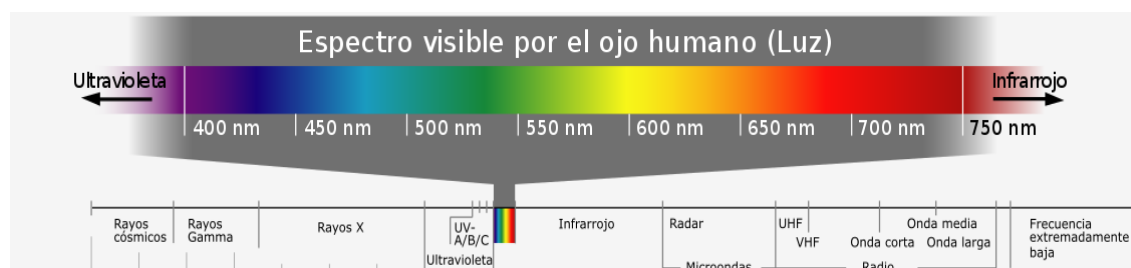


Figura 2: Espectro visible por el ojo humano

Fuente: Biología Cotidiana



Tabla 9

Límite de la luz visible según los colores RGB

Colores	LIMITES
Azul	0.4 - 0.5 μm
Verde	0.5 - 0.6 μm
Rojo	0.6 - 0.7 μm

Fuente: Introduction to Remote Sensing Fift Edition

1.3.4. Radiancia

Se le llama radiancia, a la radiación del sol, que cuyo trayecto proviene dado de la relación superficie - atmosfera - sensor. Esta magnitud es la más fundamental en la teledetección ya que el sensor lo que guarda, es la radiancia al nivel de sensor transformada a niveles digitales que dependerán de la resolución radiométrica del mismo (Frulla, 1992).

1.3.5. Radiancia directa

Es la que entra a los sensores sin extravió, con la correspondencia a la dirección inicial, llamada igualmente radiancia del pixel (Frulla, 1992).

Radiancia directa es procedente de la atmosfera.

Se hace mención aquella luz que se encuentra atrapada en la atmosfera, que viene con información agregada que se considera como error en la señal obtenida por el sensor (Frulla, 1992).



1.3.6. Radiancia proveniente del entorno

Es la radiación que se emite por sectores vecinas al pixel observado en un santiamén de tiempo, varios algoritmos de corrección atmosférica corrigen este efecto por medio de filtros (Frulla, 1992).

Ecuación de la Radiancia

$$L = \frac{\partial I}{\partial A * \cos \theta}$$

Donde:

L = Radiancia

∂I = Intensidad de la radiancia

∂A = Unidad de área

$\cos \theta$ = coseno de la dirección

Por lo tanto “la radiancia espectral (L_T) es la energía radiada en una determinada longitud de onda” (Pérez & Muñoz, 2006)

$$L_T = \frac{\partial L}{\partial \lambda}$$

Donde:

L_T = Radiancia espectral

∂L = Energía radiada

$\partial \lambda$ = Longitud de onda



Radiancia

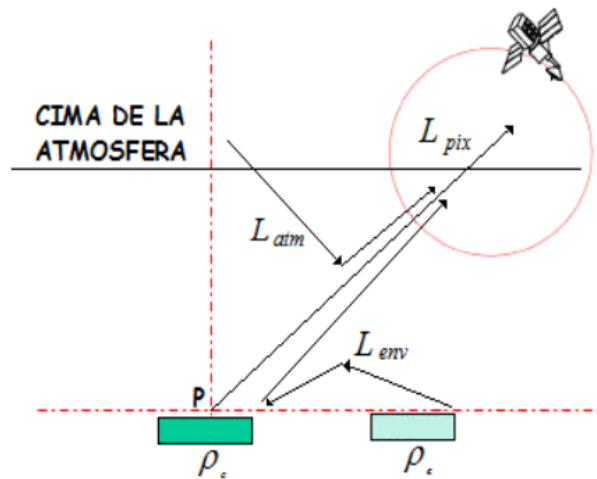


Figura 3: Componente De La Radiancia

Fuente: Detección de la Salinidad en los Suelos de Cultivos de Algodón usando Imágenes de Satélite en la Costa Norte del Perú

1.3.7. Reflectancia

Es la relación entre la radiación reflejada y el incidente, cuando se habló de correcciones atmosféricas se considera dos tipos de reflectancia, la reflectancia en la cima de la atmosfera (TOA), aún conocida como reflectancia planetaria y la reflectancia a nivel de superficie (Rahman y Dedieu, 1994). La reflectancia planetaria es la relación de la radiación solar reflejada por el sistema tierra-Atmosfera a la radiación solar incidente.

Mostramos la reflectancia en la siguiente ecuación

$$p_{\lambda} = \frac{ER(\lambda)}{EI(\lambda)}$$

Donde:

p_{λ} = Reflectancia espectral

$ER(\lambda)$ = Energía reflejada

$EI(\lambda)$ = Energía incidente



1.3.8. Teledetección

Su objetivo es reconocer las propiedades del suelo terrestre y los fenómenos que se realizan a partir de los datos guardados por el sensor. Los que trabajan en el campo de teledetección su principal reto es transformar los datos guardados por el sensor en magnitudes físicas, y también brindar información apropiada al usuario (Badaracco Meza, 2012).

1.3.8.1. Imágenes multiespectrales

Es una imagen capturada mediante sensores fotoeléctricos que convierten los niveles de radiancia procedentes de la superficie terrestre en valores numéricos que se denominan números digitales para cada una de las bandas, realizando una matriz tridimensional donde cada valor de nivel digital se ubica en una fila y una columna, y en una banda del espectro electromagnético. Existen 5 tipos de resoluciones para las imágenes multiespectrales, la resolución espacial, se refiere a la cantidad de m^2 que representa cada pixel; la resolución radiométrica, representa los valores de radiancia que puede el sensor representar en valores digitales; la resolución espectral, es el número y anchura de las bandas espectrales obtenidas por el sensor; la resolución temporal, es el tiempo en el cual el satélite paso sobre una zona para tomar la imagen; y la resolución angular, la capacidad que tiene el sensor para tomar imágenes oblicuas. (Pérez & Muñoz, 2006, p. 8)

Tabla 10

Resolución espacial de satélites

SATÉLITE	SENSOR	RESOLUCION ESPACIAL
IKONOS		1 m
ORBVIEW 3		1 m
SPOT 5		2.5 m
SPOT 3	HRV	10 m
LANDSAT 7	ETM+	15 m
LANDSAT 8	OLI	30 m

Nota: m = Metros

Fuente: Teledetección Nociones y Aplicaciones



1.3.8.2. Pre procesamiento de imágenes multiespectrales de satélite

Para usar la información de las imágenes multiespectrales, es necesario hacer dos tipos de correcciones, una corrección geomática para cambiar el sistema de coordenadas y la proyección cartográfica, y una corrección radiométrica para cambiar los valores de niveles digitales de la imagen, limpiar cualquier defecto. Además, se debe realizar las calibraciones geométricas y radiométrica. La calibración geométrica limpia las distorsiones espaciales de la imagen y la calibración radiométrica transforma los niveles digitales de la imagen a valor físicos de radiancia y reflectancia (Pérez & Muñoz, 2006).

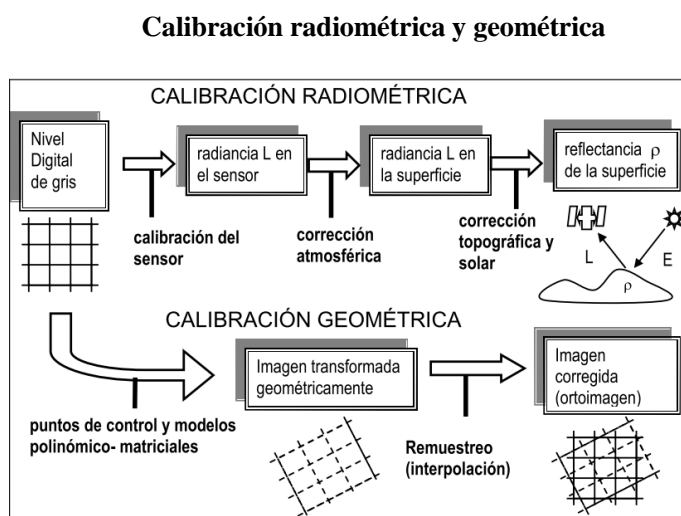


Figura 4: Proceso de calibración radiométrica y geométrica de imágenes
Fuente: Teledetección Nociones y Aplicaciones

1.3.9. Métricas de medición de técnicas.

Las siguientes métricas son utilizadas para la evaluación de técnicas y comparación entre ellas (NEGRÓN BAEZ, 2014). Las técnicas serán evaluadas con las siguientes métricas:



a) Error cuadrático medio o RMSE

La raíz del error cuadrático medio o Root Mean Square Error realiza una diferencia entre valores estimados y valores reales, la diferencia se eleva al cuadrado y se calculan el promedio de todas ellas. Este promedio se le calcula su raíz cuadrada. RMSE mide la magnitud del error (NEGRÓN BAEZ, 2014), su ecuación es la siguiente:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

Donde:

y = es el valor real

\hat{y} = es el valor estimado

n = es el tamaño de la muestra

b) Coeficiente de Determinación o R^2

El coeficiente de determinación, para esta métrica se espera un valor cercano a 1, indica que tan ajustados son los resultados obtenidos a la línea de regresión de los resultados reales. Si el resultado es 1 los datos se ajustan perfectamente (NEGRÓN BAEZ, 2014). Su ecuación es la siguiente:

$$R^2 = \frac{VE}{VT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Donde:

y = es el valor real

\hat{y} = es el valor estimado

n = es el tamaño de la muestra

\bar{y} = es el promedio de todos los valores de y



1.3.10. Técnicas de estimación

Mediante un conjunto de características numéricas que se utiliza como datos de entrada, y un resultado numérico, como variable de salida, autoriza determinar una relación entre las características y el resultado que se desea obtener, mediante la creación de relaciones y asociaciones entre las características y el resultado, con la finalidad de predecir resultados con un conjunto de características diferentes. (Sarkar, Bali, and Sharma 2018)

a) Regresión lineal múltiple

La regresión es una técnica estadística que calcula dicha similitud en función matemática. Hay bastantes técnicas de regresión en función del tipo de variables y de forma funcional supuesta entre ellas. Las más primordiales y las más eficaces en lograr tener mayor información son las lineales. La regresión lineal es el vínculo entre dos variables que tienen un perfil lineal. La regresión lineal tiene dos versiones, una es la versión SIMPLE que asocia dos variables, y la otra es la MULTIPLE que influyen más de dos variables. La fórmula de la regresión lineal múltiple es la siguiente (Montenegro Granados, 2016):

$$y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + u_j$$

Donde:

y = es la variable endógena,

x = las variables exógenas,

u = los residuos

b = los coeficientes estimados del efecto marginal entre cada x e y .



La “y” puede nombrarse como endógena, dependiente, variable respuesta, la “x” son variables que se pueden nombrar como exógena, independiente. En regresión lineal múltiple solo contamos con una variable endógena (y) y puede tener muchas variables exógenas “x” (Montenegro Granados, 2016).

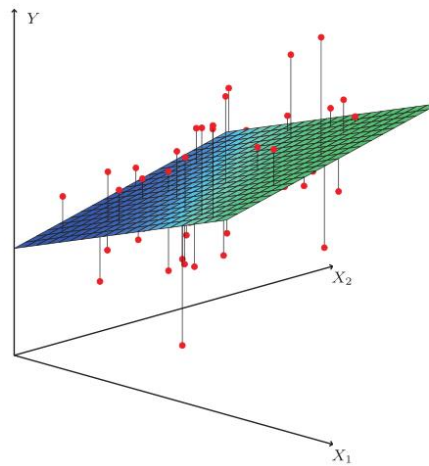


Figura 5: Regresión Lineal Múltiple
Fuente: Introducción Regresión Lineal Múltiple



b) Random Forest

Los Bosques Aleatorios es la unión de factores predictivos de árboles, de tal manera que cada árbol necesita de los valores de un vector aleatorio, mostrado de forma independiente y con la misma asignación para todos los árboles en el bosque. El error de generalización para bosques dirige hasta un límite a medida que aumenta la cantidad de “árboles” en el “bosque”. El error de generalización de un bosque de clasificadores de árboles necesita de la fuerza de los árboles individuales en el bosque y la relación entre ellos. El uso de una selección aleatoria de características para dividir cada nodo produce tasas de error que se comparan favorablemente para Adaboost (meta-algoritmo de aprendizaje automático), pero son más robustos con respecto al ruido. Las estimaciones internas monitorean el error, la fuerza y la correlación, y se emplean para exhibir la respuesta al incremento del número de características empleadas en la división. Las estimaciones internas se pueden utilizar para medir la importancia de las variables. Estas ideas también son aplicables a la regresión (Breiman, 2001).

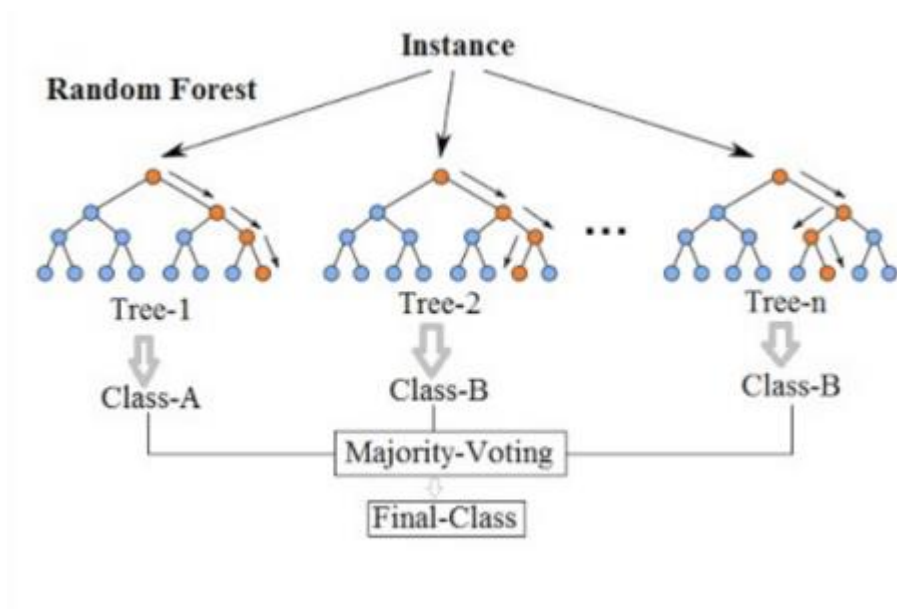


Figura 6: Random Forest

Fuente: Machine Learning Basics – Random Forest



1.4. Formulación del problema

¿Qué técnica tendrá mejor rendimiento al estimar la salinidad en suelos con contenido de humedad?

1.5. Justificación e importancia del problema

El uso de la tecnología del dominio agrícola o cualquier ámbito de la actividad humana se ha vuelto indispensable debido a la rapidez del procesamiento de la información y bajo costo, las imágenes multiespectrales son una herramienta al cual se puede tener acceso gratuito y luego procesarlas con técnicas de ingeniería, se puede sacar provecho para mejorar la productividad agropecuaria en la región Lambayeque.

En el aspecto económico, la detección de salinidad con los métodos tradicionales en un terreno demasiado grande, sería muy costoso, por lo cual utilizaremos las imágenes multiespectrales para detectar la salinidad, ya que sería un método menos costoso para el agricultor.

En el aspecto ecológico, se justifica que en el departamento de Lambayeque se ha visto bastante pérdida de cosechas, ya que esto se debe al mal drenaje de agua y a los suelos que tienen demasiada salinidad, es por ello que posiblemente se reduzca la pérdida de cosechas.

En el aspecto social, se justifica que la detección de salinidad en los suelos, mejorara las cosechas y el bienestar del agricultor.

Como podemos apreciar, la importancia de este trabajo de investigación es obtener la técnica de estimación de salinidad más preciso por medio de imágenes multiespectrales, para que así poder mejorar la productividad agropecuaria.

1.6. Hipótesis

Utilizando la teledetección en el procesamiento de imágenes multiespectrales, y la aplicación de las técnicas de estimación con Random Forest y Regresión Lineal Múltiple, se podrá estimar el grado de salinidad en suelos con contenido de humedad.

La estimación del grado de salinidad con Random Forest, obtendrá mejores resultados en los indicadores usados (R^2 y **RMSE**), que la técnica de Regresión Lineal Múltiple, por lo tanto, al evaluar ambas técnicas en dichos indicadores (**Anexo 1**), la precisión al utilizar Random Forest es mayor a la precisión de Regresión Lineal Múltiple.

1.7. Objetivos

1.7.1. Objetivo General

Comparar técnicas de estimación del grado de salinidad en imágenes multiespectrales en suelos con contenido de humedad.

1.7.2. Objetivo Específicos

1. Seleccionar las técnicas de estimación a utilizar.
2. Adquisición de imágenes multiespectrales del área de estudio
3. Realizar calibraciones atmosféricas a las imágenes multiespectrales adquiridas.
4. Seleccionar indicadores de salinidad.
5. Recolectar muestras de suelos con contenido de humedad y realizar la prueba de conductividad eléctrica.
6. Evaluar las técnicas de estimación.



II. MATERIAL Y METODOS

2.1. Tipo y diseño de investigación

2.1.1. Tipo de investigación.

La investigación Cuantitativa es donde se recolecta, procesan y examinan datos cuantitativos o numéricos sobre variables, también estudia la asociación entre variables cuantificadas donde nos ayuda más en la interpretación de los resultados. (Yanetsys Sarduy, 2006).

2.1.2. Diseño de investigación.

La investigación cuasi experimental estudia las asociaciones de causa-efecto entre las variables independientes y dependientes (Agudelo et al.,2008).

2.2. Población y muestra

La población son las técnicas de estimación mediante información proveniente de imágenes multiespectrales (Anexo 1) y la muestra escogida son las técnicas de estimación de Random Forest y Regresión Lineal Múltiple.

2.3. Variables

2.3.1. Variable Independiente

Técnica de estimación.

2.3.2. Variable Dependiente

Grado de salinidad en suelos con contenido de humedad.



2.3.3. Operacionalización

Variable Independiente	Dimensiones	Indicadores	Unidad de Medida	Formula	Descripción
Técnicas de estimación de grado de salinidad	Rendimiento de los valores de salinidad estimados	Coficiente de determinación	Porcentaje	$R^2 = \frac{VE}{VT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	y = es el valor real \hat{y} = es el valor estimado n = es el tamaño de la muestra \bar{y} = es el promedio de todos los valores de y
		Raíz del Error cuadrático medio	Porcentaje	$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$	y = es el valor real \hat{y} = es el valor estimado n = es el tamaño de la muestra

Variable Dependiente	Dimensiones	Indicadores	Unidad de Medida	Formula	Descripción
Grado de salinidad en suelos	Grado de Salinidad	Conductividad Eléctrica (CE)	ds/m (deciSiemens / metro)	$CE_{25} = CE_t \times ft$	CE_t = La conductividad eléctrica medida del extracto de tierra en t °C (grados centígrados) ft = es el valor corregido de la conductividad eléctrica en t °C (grados centígrados)



2.4. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.

2.4.1. Abordaje Metodológico.

El abordaje metodológico fue cuasi experimental, ya que este tipo de diseño estudia las asociaciones de causa-efecto entre las variables independientes y dependientes (**Agudelo et al.,2008**).

2.4.2. Técnicas de Recolección de datos.

Se aplicó la observación científica, ya que se utilizarán imágenes multiespectrales para realizar el proyecto de investigación.

Se recolecto documentos como guías, artículos de investigación, papers y libros.

2.5. Procedimientos de análisis de datos

Una vez identificados los métodos de estimación a utilizar se realizará lo siguiente:

- a) En primera fase se debe obtener las imágenes multiespectrales del área de estudio, las cuales serán descargadas gratuitamente de la página de Landsat.
- b) En segunda fase se procede a realizar las calibraciones atmosféricas para obtener la reflectancia por pixel de la imagen.
- c) En tercera fase se segmentará la imagen a la zona de estudio.
- d) En cuarta fase se calcularán los índices espectrales de salinidad.
- e) En quinta fase se estimará la salinidad con Regresión Lineal Múltiple y Random Forest.

2.5.1. Análisis estadístico de datos

Para examinar y explicar los resultados obtenidos de los indicadores en la operacionalización de variables, se realizará un procesamiento estadístico en lenguaje Python utilizando la librería scikit-learn.



2.6. Criterios éticos

En el trabajo de investigación se aplicó principios éticos de honestidad y solidaridad, ya que se citó de manera correcta y se referencio los libros, papers, que se utilizaron para realizar el trabajo de investigación.

2.7. Principios de rigor científico

Fiabilidad: La investigación cumplió con las expectativas fijadas en su contenido y su implementación.

Validez: Los datos obtenidos en la investigación fueron evaluados y analizados por personal especializado en el tema para decretar su autenticidad.

2.8. Método Propuesto



Figura 7: Método Propuesto de la tesis “Comparación De Técnicas De Estimación Del Grado De Salinidad En Suelos Con Contenido De Humedad Mediante El Procesamiento De Imágenes Multiespectrales”

Fuente: Elaboración Propia

2.8.1. Seleccionar las técnicas de estimación a utilizar.

La población de técnicas consideradas para esta investigación se lista en el **Anexo 1**, se eligió 2 técnicas de estimación con los siguientes criterios.

Según el trabajo de investigación “The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data” (Kayri et al., 2017), nos indica que Random Forest tiene un coeficiente de correlación de un porcentaje de 98.60% y la regresión lineal múltiple tiene el coeficiente de correlación con un porcentaje de 96.30% de precisión.

Según el trabajo de investigación “ A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy” (Yue et al., 2018), nos indica que Random Forest tiene un coeficiente de correlación de un porcentaje de 85.00% y la regresión lineal múltiple tiene el coeficiente de correlación con un porcentaje de 87.00% de precisión.

Con base a los trabajos anteriores se propuso comparar la técnica Random Forest, que obtuvo mejores resultados que la técnica de Regresión Lineal Múltiple, usada en los trabajos de investigación de estimación de salinidad usando imágenes multiespectrales, además posee un estudio previo en salinidad.

Por tal motivo se escogió ambas técnicas para aplicarlas y saber cual es la mejor para esta investigación.

2.8.2. Adquisición de imágenes multiespectrales del área de estudio.

El objetivo principal de este proyecto de investigación es comparar las técnicas de estimación del nivel de salinidad que se realizaran en áreas de terrenos con contenido de humedad, se requiere la información y una base de datos de las imágenes multiespectrales del área de estudio. Dicha información es capturada mediante el satélite Landsat 8 y a que no se encontró una base de datos ya elaborada, lista para procesar, se realizó una base de datos con imágenes multiespectrales con el área indicada.



El satélite LANDSAT-8 fue lanzado el 11 febrero de 2013, es un satélite óptico de resolución media cuya finalidad es brindar información para actividades con la educación, la agricultura y para el ámbito estatal. Contiene dos sensores:

Sensor operacional de imágenes de tierra (OLI), está integrado por 9 bandas espectrales que abarca desde lo visible, transitando por los infrarrojos, hasta el espectro de radiación de bajas ondas. Dando así imágenes de resolución espaciales, que oscilan entre los 15m y los 30m.

Sensor Infrarrojos Térmicos (TIRS), está formado por dos bandas espectrales que localizan infrarrojos térmicos, una herramienta para medir la temperatura de la superficie de la tierra. (Geocento, 2018)

En la tabla se muestra las 7 primeras bandas espectrales para se utilizaron en la presente investigación.

Tabla 11

Bandas del sensor OLI del satélite Landsat 8

Bandas	Longitud de Onda (nm)	Límites de espectro electromagnético (um) micras	GSD (m) por lado
1 - COASTAL AEROSOL	300	0.435 – 0.451	30
2 – AZUL	450 - 510	0.452 – 0.512	30
3 – VERDE	530 – 590	0.533 – 0.512	30
4 – ROJO	640 - 670	0.636 – 0.673	30
5- INFRARROJOS	850 – 880	0.851 – 0.879	30
6 - SWIR 1	1570 - 1650	1.566 – 1.651	30
7 - SWIR 2	2110 – 2290	2.107 – 2.294	30

Nota: nm= nanometros; um=micrómetros; m=metros; Límites del espectro electromagnético= Rango de valores que contiene una banda electromagnética; GSD= Esta nos indica la distancia entre dos centros de pixeles consecutivos medidos en el suelo.

Fuente: Geocento earth imaging



Las capturas de las imágenes multiespectrales del Landsat 8, se lograron obtener del sitio web USGS de manera gratuita, se realizaron los siguientes pasos en el **anexo 2**, al final nos brinda una descarga de una carpeta con la siguiente estructura:

Imágenes Multiespectrales

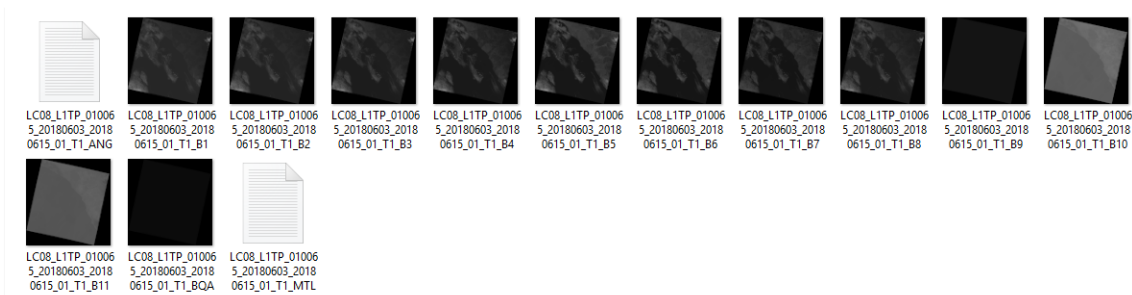


Figura 8: Imágenes Multiespectrales descargadas de Landsat 8
Fuente: Elaboración Propia

Donde solamente utilizaremos las imágenes multiespectrales que se encuentran desde la banda 1 hasta la banda 7, y los dos archivos en formato de texto (MTL.txt, ANG.txt) que incluyen en la información entregada por Landsat 8.

Información de las Imágenes multiespectrales

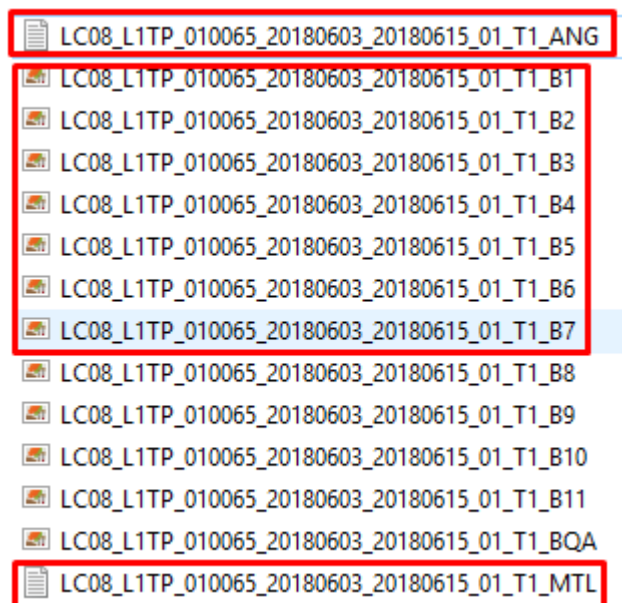


Figura 9: Información de las Imágenes Multiespectrales descargadas de Landsat 8
Fuente: Elaboración Propia



Tabla 12

Descripción de los archivos Landsat 8

MTL	ANG
<ul style="list-style-type: none"> • Tiene información útil para la búsqueda sistemática y las prácticas de archivo de datos. • Incluye información sobre el procesamiento de datos y los valores para mejorar los datos de Landsat como: la conversión a reflectancia y luminosidad. 	<ul style="list-style-type: none"> • Proporciona los coeficientes del modelo de ángulo de visión del sensor que se usan para calcular los ángulos de visión solar y sensor para cada pixel

Fuente: Elaboración Propia

Con lo mencionado anteriormente, se realizó una base de datos con todas las carpetas descargadas de Landsat 8, y ordenando por el rango de fechas en que se capturaron ciertas imágenes multiespectrales (**Anexo 2**)

Las imágenes multiespectrales del Landsat 8 requeridas por el sensor (OLI) y (TIRS) consiste de nueve (9) bandas espectrales. Por lo cual seleccionamos desde la banda 1 hasta la banda 7 por lo que se obtiene una resolución espacial de 30 metros, es decir que cada pixel equivale a $30m^2$, la banda azul profundo (1) es utilizado para estudios costeros y aerosoles. El tamaño de la imagen capturada por Landsat 8 es de 170km de norte-sur por 183km a este-oeste (106km por 114km).

2.8.3. Realizar calibraciones atmosféricas a las imágenes multiespectrales adquiridas.

Luego de adquirir las imágenes multiespectrales, se realizó la calibración atmosférica, donde se cambió los números digitales a reflectancia, y después a reflectancia superior atmosférica (TOA) utilizando variables proporcionadas en el archivo de metadatos (MLT), dicho archivo se descargó junto con las imágenes de Landsat 8. En las ecuaciones siguientes se realiza la conversión de valor digital a radiancia y reflectancia (Department of the Interior U.S., 2018).



Conversión de pixeles números digitales a radiancia.

$$L_{\lambda} = M_L Q_{cal} + A_L$$

Donde:

L_{λ} = Radiancia espectral al nivel de la superficie

M_L = Factor de re escalado multiplicativo presente en la metadata **RADIANCE_MULT_BAND_x**

A_L = Factor de re escalado aditivo presente en la metadata **RADIANCE_ADD_BAND_x**

Q_{cal} = El valor digital (ND) del pixel

Conversión de pixeles de valores digitales a reflectancia a nivel de la superficie.

$$p_{\lambda'} = M_p Q_{cal} + A_p$$

Donde:

$p_{\lambda'}$ = Radiancia espectral al nivel de la superficie

M_p = Factor de re escalado multiplicativo presente en la metadata **REFLECTANCE_MULT_BAND_x**

A_p = Factor de re escalado aditivo presente en la metadata **REFLECTANCE_ADD_BAND_x**

Q_{cal} = El valor digital (ND) del pixel

Corrección de reflectancia respecto el ángulo solar

$$p_{\lambda} = \frac{p_{\lambda'}}{\cos(\theta_{sz})} = \frac{p_{\lambda'}}{\sin(\theta_{se})}$$

$p_{\lambda'}$ = Reflectancia a nivel de la superficie

θ_{se} = Angulo solar de elevación

θ_{sz} = Angulo solar zenith



Para este proceso de investigación se desarrolló la transformación de valores digitales a reflectancia con corrección angular, para eso se transformó cada pixel de cada una de las imágenes (**anexo 3**), se programó un pequeño algoritmo en Python que ejecuta el proceso automáticamente (**anexo 4**).

Se obtuvieron siete imágenes nuevas que se guardan en memoria del programa, y estos contienen los valores de reflectancia de cada imagen espectral original.

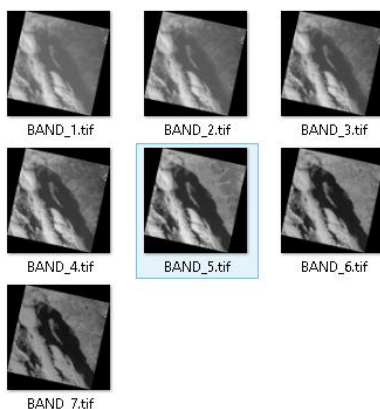


Figura 10: Imágenes De Reflectancia Generadas Al Procesar Las Imágenes
Fuente: Elaboración Propia

Estas imágenes podemos identificar en el siguiente cuadro:

Tabla 13

Imágenes espectrales de reflectancia, banda y su identificador

IDENTIFICADOR	IMAGEN	BANDA
p_B	BAND_2	Banda Azul
p_G	BAND_3	Banda Verde
p_R	BAND_4	Banda Roja
p_{NIR}	BAND_5	Banda infrarrojo cercano
p_{SWIR}	BAND_6	Banda infrarrojo de onda corta

Fuente: Elaboración propia.



2.8.4. Seleccionar indicadores de salinidad

Luego de haber la realizado la conversión de cada pixel de imagen a la reflectancia superior atmosférica (TOA) se ejecutaron los indicadores de salinidad y bandas, se escogió los siguientes indicadores en los trabajos de investigación:

En el trabajo de investigación de “**Digital Mapping of Topsoil Salinity Using Remote Sensing Indices in Agh-Ghala Plain, Iran**” (Mousavi et al., 2017), realizaron la comparación de 17 indicadores de salinidad y de bandas espectrales con la finalidad hallar la correlación de ciertos indicadores y el valor de conductividad eléctrica.

Tabla 14

Correlación entre los indicadores de salinidad, vegetación y bandas, y el valor de conductividad eléctrica

INDICADOR	Coefficiente de correlación (R ²)
ASTER-SI	0.20 - 0.40
SI1	0.20 - 0.40
SI2	0.00 - 0.20
SI3	0.20 - 0.40
SI11	0.20 - 0.40
INT1	0.00 - 0.20
B2 (Banda azul)	0.40 - 0.50
B3 (Banda verde)	0.30 - 0.40
B4 (Banda Roja)	0.20 - 0.30

Fuente: Digital Mapping of Topsoil Salinity Using Remote Sensing Indices in Agh-Ghala Plain, Iran.

Para realizar el trabajo de investigación se escogieron los siguiente indicadores ASTER SI, SI1, SI2, SI3, SI11, INT1, B2, B3, B4 por tener una correlación mayor a 0.20.



En el trabajo de investigación “**Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments Jianli**” (Ding & Yu, 2014), se comparó cinco indicadores de salinidad para hallar la correlación entre ellos y el valor de conductividad eléctrica.

Tabla 15

Correlación entre los indicadores de salinidad, vegetación, y el valor de conductividad eléctrica

INDICADOR	Coefficiente de correlación (R ²)
SI	0.310
NDSI	0.282
BI	0.338
NDVI	0.343
DVI	0.318

Fuente: Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments Jianli.

Para realizar el trabajo de investigación escogemos los siguientes indicadores NDVI, DVI, BI, SI por tener una correlación mayor a 0.30 en relación a la conductividad eléctrica.

En el trabajo de investigación “**Mapping and Modelling Spatial Variation in Soil Salinity in the Al Hassa Oasis Based on Remote Sensing Indicators and Regression Techniques**” (Allbed et al., 2014), se utilizó un indicador de salinidad (SI).



De los puntos anteriores, los indicadores seleccionados poseen las siguientes formulas:

$$NDVI = \frac{p_{NIR} - p_R}{p_{NIR} + p_R} \quad \text{Rouse et al. (1974)}$$

$$DVI = p_{NIR} - p_R \quad \text{Richardson y Wiegand (1977)}$$

$$SI 2 = \frac{p_R}{p_{NIR}} \quad \text{Tripathi et al. (1997)}$$

$$BI = \sqrt{p_G^2 + p_R^2} \quad \text{Inakwu and Odeh (2008)}$$

$$SI = \sqrt{p_B * p_R^2} \quad \text{Khan et al. (2005)}$$

$$SI 11 = \frac{p_{NIR}}{p_{SWIR}} \quad \text{Khan et al. (2005)}$$

$$SI 1 = \sqrt{p_G * p_R} \quad \text{Dehni and Lounis (2012)}$$

$$ASTER SI = \frac{p_{NIR} - p_{SWIR}}{p_{NIR} + p_{SWIR}} \quad \text{Asfaw et al. (2016)}$$

$$SI 3 = \frac{p_G \times p_R}{p_B} \quad \text{Elhag (2016)}$$

Donde:

- p_B = Reflectancia en la banda azul
- p_G = Reflectancia en la banda verde
- p_R = Reflectancia en la banda roja
- p_{NIR} = Reflectancia en la banda infrarrojo cercano
- p_{SWIR} = Reflectancia en la banda infrarroja de onda corta

Para resolver cada indicador se tuvo que procesar los pixeles de cada imagen, de acuerdo a su fórmula matemática (**anexo 5**) además se realizó un pequeño algoritmo que realice este proceso automáticamente **anexo 6**.



Como resultado se crearon imágenes con los valores de cada indicador.

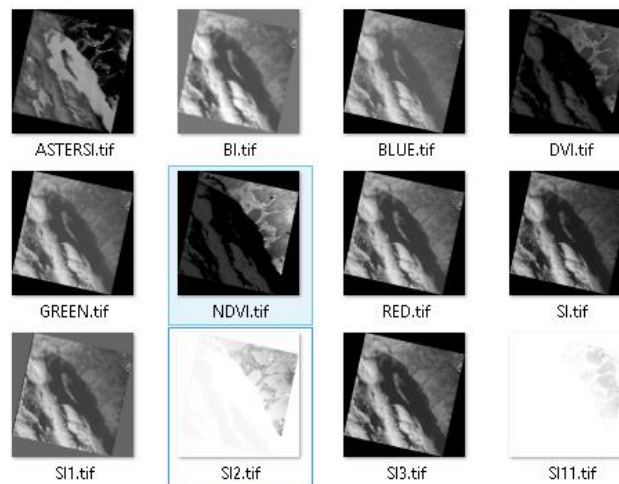


Figura 11: Imágenes Procesadas Con Los Indicadores Antes Mencionados
Fuente: Elaboración Propia

2.8.5. Recolectar muestras de suelos con contenido de humedad y realizar la prueba de conductividad eléctrica

Para este proyecto de investigación se escogió el distrito de Morrope, Lambayeque para recopilar ciertas muestras de tierra, y así mismo realizar un análisis de pruebas de conductividad eléctrica CE, ya que gracias a la CE nos sirve para medir la concentración de sales en una solución a partir de un extracto de solución de suelo.

Para este análisis se tomaron 10 muestras de tierra y a la vez 10 puntos de muestreo a una profundidad de 30 cm del suelo, geo referenciados usando un sistema de Posicionamiento Global (GPS), con un sistema de coordenadas UTM-WSG84, zona 17S.

La distancia entre las muestras fue de mínimo 900 m², la misma que tiene cada pixel en las imágenes multiespectrales proporcionadas por el satélite Landsat 8, con el fin de obtener muestras de diferentes pixeles en suelos de escasa vegetación, la zona también tenía suelo de abundante vegetación a los alrededores por tal motivo se tomó otras 4 muestras más lejanas.

El GPS utilizado fue un Garmin Montana 650t el cual tiene una precisión entre +- 3.65m. (“Montana 600 Series Owner’s Manual,” 2011), dicha precisión es suficiente debido al tamaño de pixel antes mencionado.



El procedimiento que se realizó a las muestras fue el siguiente:

1. Se realizó el secado de las muestras de tierra, extendiéndolas al aire libre sobre un depósito.
2. Se procedió a moler la muestra con un rodillo para deshacer los agregados del suelo
3. Se tamizó la muestra de suelo con el tamiz N° 10, separando los elementos gruesos de la tierra fina, y se pesó y agregó 500 gramos de suelo tamizado en un recipiente de plástico.
4. Se agregó agua destilada al recipiente de plástico, y se bate con una espátula hasta llegar a una pasta de saturación.
5. En un embudo se colocó un papel filtro y se agregó la pasta saturada, y con una bomba al vacío se extrajo el extracto del suelo
6. Al final se calcula la conductividad eléctrica del extracto del suelo, con un conductímetro.

Tabla 16

Ubicación de las muestras tomadas y su respectivo valor de conductividad eléctrica

N° PRUEBA	UBICACIÓN UTM 17S	VALOR DE CONDUCTIVIDAD ELECTRICA (CE)
A1	617218, 9265607	0.9
A2	617211, 9265623	1.22
A3	617183, 9265653	0.84
A4	617156, 9265671	1.50
A5	617153, 9265644	1.60
A6	617140, 9265616	1.0
A7	617208, 9265589	1.4
A8	617196, 9265572	1.78
A9	617187, 9265579	1.6
A10	617171, 9265583	1.46

Fuente: Elaboración Propia



Luego de obtener las coordenadas de las muestras TOMADAS por GPS, se realizó la ubicación de cada muestra en el formato UTM 17S EN LAS imágenes multiespectrales (tabla 14), es por ello que las coordenadas en UTM 17S se tendrá que convertir a UTM17N porque este es un estándar del Landsat 8.

Para extraer los indicadores se toman como guía dicho proceso que se encuentra en el (ANEXO 7), y se desarrolló un programa que realiza este proceso automáticamente con ayuda de la librería GDAL (ANEXO 8).

En la siguiente tabla mostramos los indicadores y los valores de conductividad.

Tabla 17
Relación entre los indicadores elegidos y el valor de conductividad eléctrica

NUM.	RED	GREEN	BLUE	NDVI	DVI	BI	ASTERSI	SI	SI1	SI2	SI3	SI11	EC
1	0.171939075	0.154217556	0.152963549	0.131987572	0.052289166	0.230967745	0.062240254	0.067246333	0.162837416	0.766803861	0.173348635	0.882813215	0.9
2	0.159990653	0.147923917	0.148988634	0.233333334	0.097385615	0.217895612	0.015591448	0.061754864	0.153839022	0.621621609	0.158847317	1.031676769	1.22
3	0.142529383	0.140636563	0.14385435	0.307387173	0.126511395	0.200233027	0.024322134	0.054058738	0.141579807	0.529768705	0.13934122	1.049856901	0.84
4	0.162427664	0.152443036	0.151307344	0.255140245	0.111274183	0.222759113	0.019476514	0.063181505	0.157356173	0.593447447	0.163646817	1.039726734	1.5
5	0.171962738	0.157033116	0.152750611	0.204509392	0.088418379	0.232874617	0.015873017	0.067208752	0.164328471	0.660427094	0.176783875	0.96875	1.6
6	0.183627248	0.165148586	0.158831298	0.170079663	0.075263277	0.246967658	0.024733722	0.073182151	0.174142987	0.709285319	0.190930754	0.951726556	1
7	0.171939075	0.154217556	0.152963549	0.131987572	0.052289166	0.230967745	0.062240254	0.067246333	0.162837416	0.766803861	0.173348635	0.882813215	1.4
8	0.174801975	0.156370625	0.155116633	0.130516648	0.052478448	0.234536782	0.057496075	0.068845496	0.16532965	0.769102633	0.176215112	0.891259968	1.78
9	0.178232715	0.158712998	0.155566186	0.125544146	0.051177133	0.238656059	0.063324153	0.070298329	0.168189913	0.776918292	0.181838021	0.880894005	1.6
10	0.187744126	0.165195912	0.160534844	0.130077288	0.056145787	0.250075072	0.062781289	0.075223066	0.176109523	0.769790471	0.193195209	0.881854713	1.46

Fuente: Elaboración Propia.



2.8.6. Evaluar las técnicas de estimación.

En este proyecto de investigación se empleó la información de las relaciones entre los indicadores y bandas (valores de entrada) y el valor de conductividad eléctrica (valor de salida) para estimar la salinidad, además se ejecutaron las técnicas Regresión Lineal Múltiple y Random Forest, los indicadores que se usaron para medir el rendimiento de ambas técnicas fueron el coeficiente de determinación (R^2) y la raíz del error cuadrático medio ($RMSE$).

Las técnicas se implantaron en Python y usando la librería scikit.

Técnica de la Regresión Lineal Múltiple

De toda la matriz de mis indicadores, con la ayuda del Excel hallamos los coeficientes (B) (Beta)

Tabla 18

Coefficientes

	Beta	Coefficientes
Intercepción	B0	378.607322
Variable X 1	B1	1689.455832
Variable X 2	B2	0.000000
Variable X 3	B3	-1299.431191
Variable X 4	B4	472.937714
Variable X 5	B5	778.842459
Variable X 6	B6	0.000000
Variable X 7	B7	1654.254051
Variable X 8	B8	0.000000
Variable X 9	B9	0.000000
Variable X 10	B10	591.575772
Variable X 11	B11	-1024.604297
Variable X 12	B12	-844.311952

Fuente: Elaboración propia.



Aplicamos la fórmula de la regresión Lineal Múltiple.

$$y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + u_j$$

Hallamos la \hat{Y} estimada, haciendo la sumatoria de coeficientes por indicador de cada muestra, $\hat{Y} = (\text{Suma (Coeficientes*Indicadores)})$. Como resultado tenemos lo siguiente:

Tabla 19

Resultado De La (\hat{Y}) Estimada y (Y) Real

ESTIMADO	REAL
\hat{Y}	Y
	EC
1.150	0.900
1.220	1.220
0.840	0.840
1.500	1.500
1.600	1.600
1.000	1.000
1.150	1.400
1.780	1.780
1.600	1.600
1.460	1.460

Fuente: Elaboración Propia

Luego hallamos el Root Mean Squared Error (RMSE).

Para hallar el RMSE se trabaja la columna de los valores estimadas y la columna de los valores reales.

Aplicamos la formula del RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$



Hallamos RMSE haciendo la raíz cuadrada de la sumatoria de la resta de \hat{Y} (Estimada) y Y (Real) al cuadrado entre el número de Muestras, $RMSE = \text{Raíz} (\text{Suma } ((\hat{Y} - Y)^2) / \text{Muestras})$. Como resultado tenemos lo siguiente:

Tabla 20

Resultado Del RMSE

RMSE
0.06250000
0.00000000
0.00000000
0.00000000
0.00000000
0.00000000
0.06250000
0.00000000
0.00000000
0.00000000
Resultado RMSE
0.11180340

Fuente: Elaboración Propia

$RMSE = 0.11180340$

Luego hallaremos el R^2 Aquired (R^2)

Para hallar el R^2 se trabaja la columna de \hat{Y} (Estimada), Y (Real) y con el Promedio del Y (Real).

Tabla 21

Promedio del Y (Real)

Promedio Del Y (Real):	1.330
-------------------------------	--------------

Fuente: Elaboración Propia



Aplicamos la formula del R^2

$$R^2 = \frac{VE}{VT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Hallamos R^2 haciendo la división de la sumatoria de la resta de \hat{Y} (Estimada) y \bar{Y} (Promedio) al cuadrado, entre la sumatoria de la resta de Y (Real) y \bar{Y} (Promedio) al cuadrado, $R^2 = \text{División (Suma}((\hat{Y} - \bar{Y})^2) / \text{Suma}((Y - \bar{Y})^2))$. Como resultado tenemos lo siguiente:

Tabla 22

Resultado Del R^2

R^2		
0.0324	0.1849	Resultado del R^2
0.0121	0.0121	
0.2401	0.2401	
0.0289	0.0289	
0.0729	0.0729	
0.1089	0.1089	
0.0324	0.0049	
0.2025	0.2025	
0.0729	0.0729	
0.0169	0.0169	
Suma	Suma	
0.8200	0.945	0.86772487
NUMERADOR	DENOMINADOR	

Fuente: Elaboración Propia.



Tabla 23

RESULTADOS DE LA ESTIMACION USANDO REGRESIÓN LINEAL MÚLTIPLE

N°	CE Estimada	CE Real
A1	1.15	0.9
A2	1.22	1.22
A3	0.84	0.84
A4	1.5	1.5
A5	1.6	1.6
A6	1	1
A7	1.15	1.4
A8	1.78	1.78
A9	1.6	1.6
A10	1.46	1.46

Fuente: Elaboración Propia

R Aqred (R^2): 0.867724867725

Root Mean Squared Error (RMSE): 0.111803398875

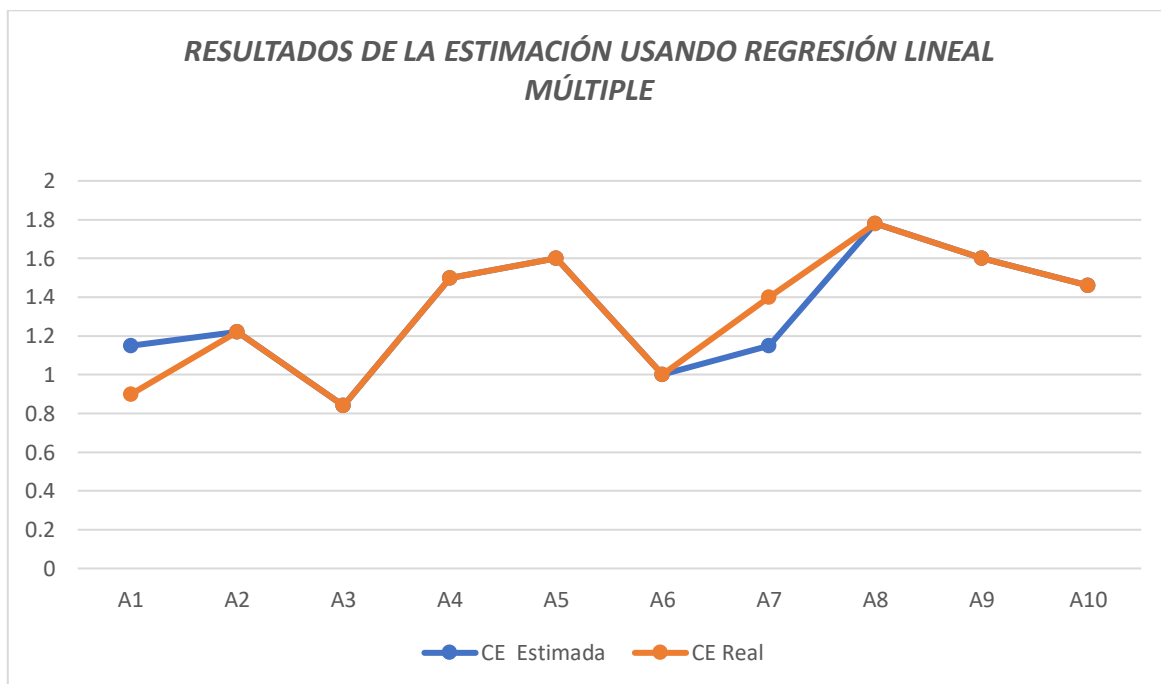


Figura 12: Resultados De La Estimación Usando Regresión Lineal Múltiple

Fuente: Elaboración Propia



Tabla 24

RESULTADOS DE LA ESTIMACION USANDO RANDOM FOREST

N°	CE Estimada	CE Real
A1	1.21	0.9
A2	1.28	1.22
A3	0.99	0.84
A4	1.39	1.5
A5	1.46	1.6
A6	1.17	1
A7	1.21	1.4
A8	1.62	1.78
A9	1.56	1.6
A10	1.43	1.46

Fuente: Elaboración Propia

R Squared (R^2): 0.73629008622

Root Mean Squared Error (RMSE): 0.157862556841

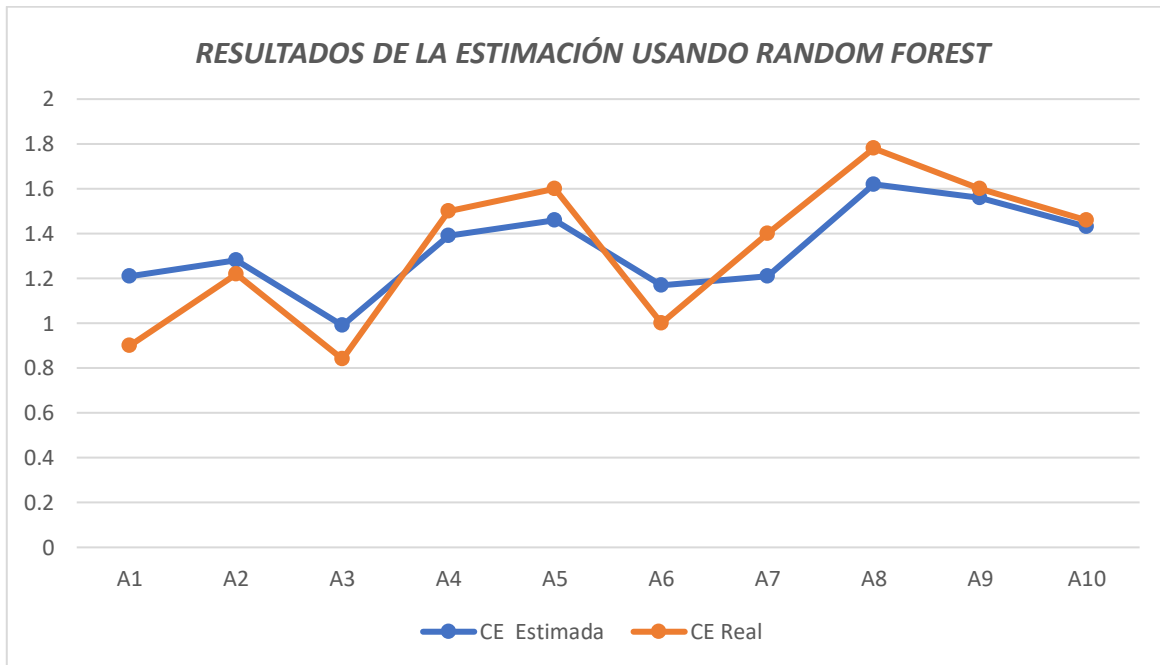


Figura 13: Resultados De La Estimación Usando Random Forest

Fuente: Elaboración Propia



III. RESULTADOS

3.1. Resultados en tablas y figuras

La hipótesis inicial para este proyecto de investigación, fue que Random Forest obtendría mejores resultados, en base a la investigación de la cual se extrajo la población (**Anexo 1**), pero los resultados fueron diferentes; Regresión Lineal Múltiple fue la técnica que obtuvo mejores resultados en la estimación de salinidad, verificado por los indicadores de R^2 y RMSE.

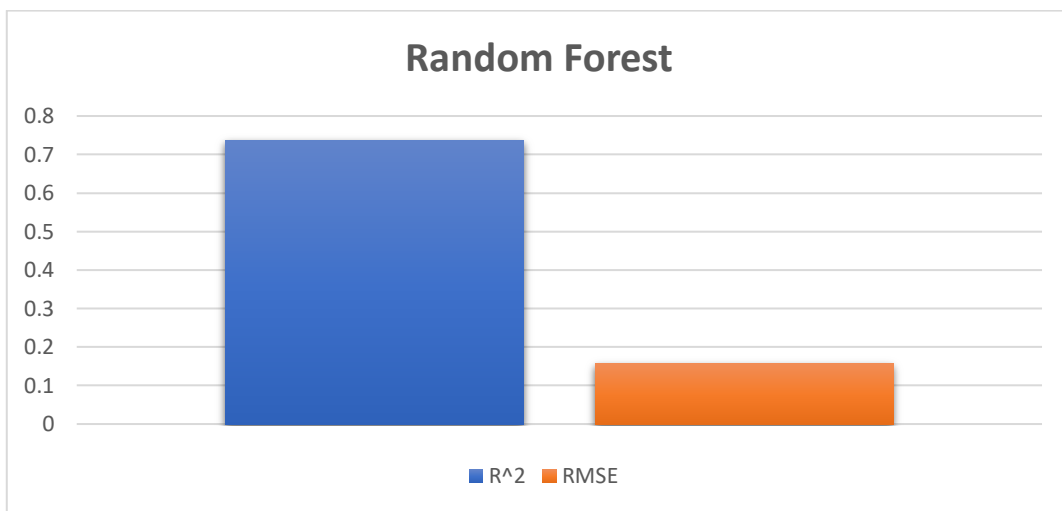


Figura 14: Resultado Del (R^2 Y RMSE) del Random Forest
Fuente: Elaboración Propia

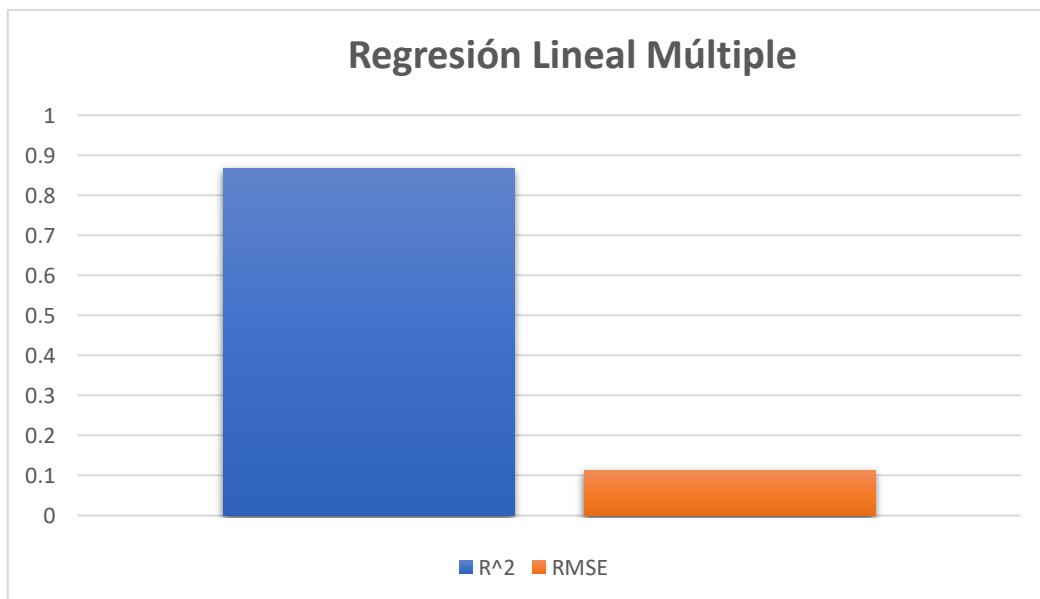


Figura 15: Resultado Del (R^2 Y RMSE) de la Regresión Lineal Múltiple
Fuente: Elaboración Propia



En el indicador de operacionalización R^2 la técnica Regresión Lineal Múltiple obtuvo 0.86 y la técnica Random Forest 0.73, lo que significó que la técnica Regresión Lineal Múltiple tuvo un 86% de precisión y la técnica Random Forest 73%.

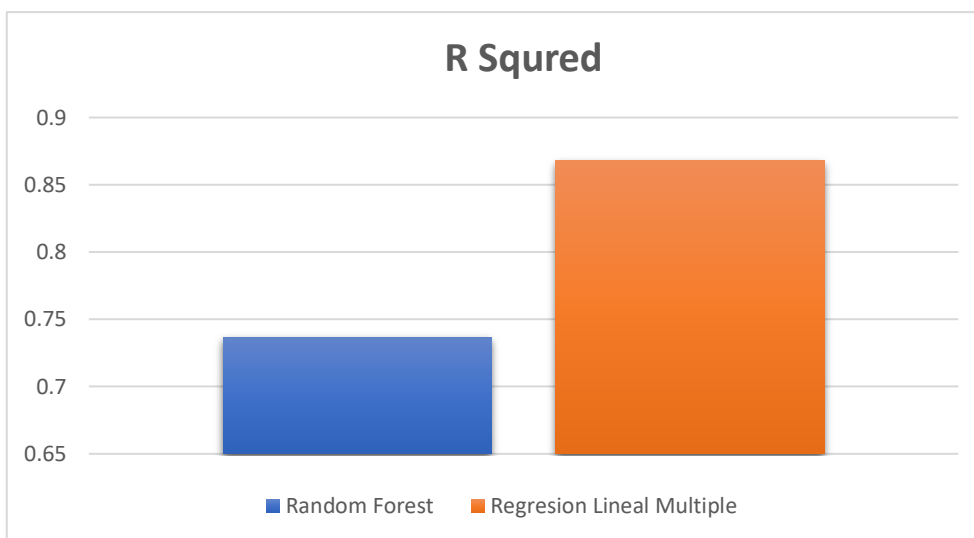


Figura 16: Resultado Del R^2 - Comparando Random Forest Y Regresión Lineal Múltiple
Fuente: Elaboración Propia

En el indicador de operacionalización RMSE la técnica Regresión Lineal Múltiple obtuvo un error de 0.11 dS/m y la técnica Random Forest 0.15 dS/m, ambas en error en medida de Conductividad Eléctrica (EC).

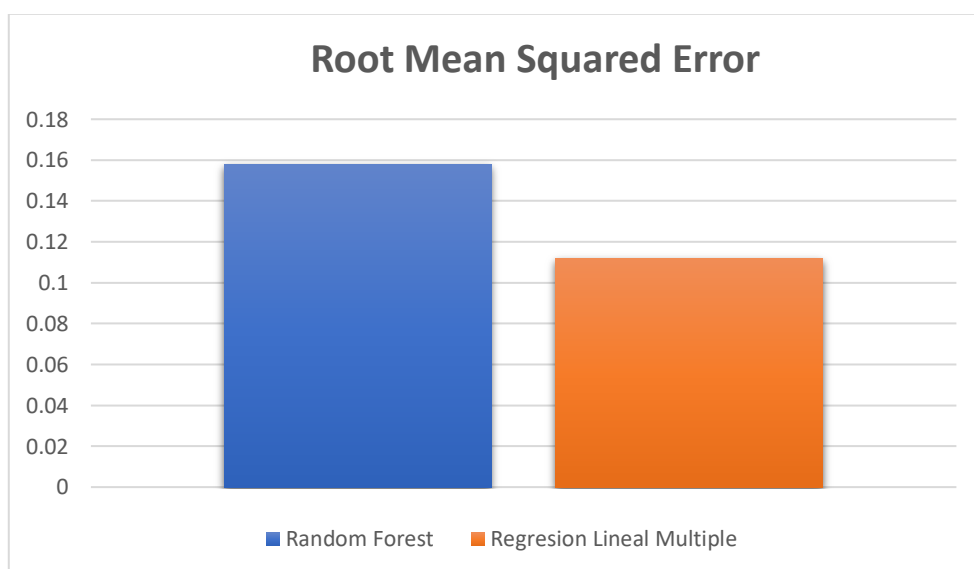


Figura 17: Resultado Del RMSE - Comparando Random Forest Y Regresión Lineal Múltiple
Fuente: Elaboración Propia



IV. CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

1. La técnica Random Forest tuvo buenos resultados en los trabajos de investigación relacionados con imágenes multiespectrales, siendo esta la más precisa al estimar la salinidad, por otra parte, la técnica Regresión Lineal Múltiple también tuvo buenos resultados al estimar la salinidad mediante propiedades físicas del ambiente, por tal motivo se escogió ambas técnicas.
2. Las imágenes multiespectrales satelitales adquiridas de la región de Lambayeque fueron gratuitas y de fácil acceso, el único problema fue la nubosidad en las imágenes, por tal motivo solo se utilizó aquellas que tenían poca nubosidad para un mejor estudio.
3. La realización de calibraciones atmosféricas a las imágenes multiespectrales, permite poder extraer indicadores de salinidad de acuerdo a las reflectancias de las bandas espectrales.
4. El criterio de selección de los indicadores de salinidad fue el porcentaje de correlación con el valor de conductividad eléctrica, una mayor relación permite análisis más precisos, y estos indicadores fueron extraídos de trabajos de investigación relacionados.
5. Para estimar la salinidad se debe tomar muestras de terreno con contenido de humedad, para la evaluación de la salinidad de las muestras se realizó la prueba de conductividad eléctrica.
6. Para evaluar las técnicas de estimación se utilizó los criterios de R^2 y RMSE, el R^2 es el porcentaje de precisión en la estimación y el RMSE es el error cuadrático en la estimación, la técnica de Random Forest obtuvo un 73% y 0.15 en cada uno respectivamente, y la técnica Regresión Lineal Múltiple 86% y 0.11, los resultados demuestran que Regresión Lineal Múltiple es más preciso y tiene menos error al estimar la salinidad que Random Forest.



4.2. Recomendaciones

1. La precisión de las técnicas de estimación varía dependiendo al trabajo de investigación en el que se aplica, por ello se recomienda no asumir que una técnica de estimación es mejor que otra, si es que no se ha probado en el mismo contexto de investigación, se debería tomar dichos resultados como referencia para cuantificar las técnicas, más no para concluir que una es mejor que otra.
2. El presente trabajo de investigación se realizó con imágenes multiespectrales satelitales, un pixel de esta imagen representa 30mx30m o 900m² de terreno, se espera que para trabajos futuros las imágenes multiespectrales que se utilicen puedan ser con un vehículo aéreo no tripulado (dron) y utilizando cámaras multiespectrales para tener una mejor estimación de la salinidad en suelos con contenido de humedad.
3. Si se adquiere imágenes multiespectrales satelitales se recomienda realizar las calibraciones atmosféricas al procesar las imágenes adquiridas
4. El porcentaje de correlación entre los indicadores seleccionados y la conductividad eléctrica vario, relacionado a los trabajos de investigación que se tomaron como base, puesto que este trabajo de investigación se realizó en otro continente y otro satélite.
5. Se recomiendo tomar muestras de conductividad eléctrica, para compararlas con valores de indicadores de salinidad ya existentes, y lograr estimar la salinidad.

REFERENCIAS

- Allbed, A., Kumar, L., & Sinha, P. (2014). Mapping and modelling spatial variation in soil salinity in the Al Hassa Oasis based on remote sensing indicators and regression techniques. *Remote Sensing*, 6(2), 1137–1157. <https://doi.org/10.3390/rs6021137>
- ALVA, C. A., VAN ALPHEN, J. G., DE LA TORRE, A., & MANRIQUE, L. (1976). *Problemas de drenaje y salinidad en la costa Peruana*.
- Badaracco Meza, R. R. (2012). Detección de la Salinidad en los Suelos de Cultivos de Algodón usando Imágenes de Satélite en la Costa Norte del Perú.
- Breiman, L. E. O. (2001). Random Forests, 5–32.
- Dasgupta, S., Hossain, M. M., Huq, M., & Wheeler, D. (2015). Climate change and soil salinity: The case of coastal Bangladesh. *Ambio*, 44(8), 815–826. <https://doi.org/10.1007/s13280-015-0681-5>
- Department of the Interior U.S. (2018). LANDSAT 8 DATA USERS HANDBOOK, 8(October), 2–4.
- Ding, J., & Yu, D. (2014). Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments. *Geoderma*, 235–236, 316–322. <https://doi.org/10.1016/j.geoderma.2014.07.028>
- Fox Llenera, E. C. (2013). *Evaluación de pérdida de suelo por salinización en la parte baja de la cuenca del jequetepeque: San Pedro de Lloc (1980-2003)*. Pontificia Universidad Católica del Perú.
- Gobierno del Estado de Baja California. (2009). *Estudio bibliografico sobre la Salinidad del valle de Mexicali*.
- Kayri, M., Kayri, I., & Tunay Gencoglu, M. (2017). The Performance Comparison of Multiple Linear Regression , Random Forest and Artificial Neural Network by using Photovoltaic and Atmospheric Data, 1–4.
- Montana 600 Series Owner’s Manual. (2011). *Garmin Ltd*.
- Montenegro Granados, R. (2016). Modelos de regresión lineal múltiple.
- Mousavi, S. Z., Habibnejad, M., Kavian, A., Solaimani, K., & Khormali, F. (2017). Digital Mapping of Topsoil Salinity Using Remote Sensing Indices in, 5(2), 1771–1786. <https://doi.org/10.18869/modares.Ecopersia.5.2.1771>
- NEGRÓN BAEZ, P. A. (2014). *REDES NEURONALES SIGMOIDAL CON ALGORITMO LM PARA PRONOSTICO DE TENDENCIA DEL PRECIO DE LAS ACCIONES DEL IPSA*. Pontificia Universidad Católica de Valparaíso.
- Nila Rekha, P., Gangadharan, R., Pillai, S. M., Ramanathan, G., & Panigrahi, A. (2012). Hyperspectral image processing to detect the soil salinity in coastal watershed.
- Otero, L., Francisco, A., Gálvez, V., Roberto, M., Inalvis, S., Martha, L., ... Luis, R. (2008). Caracterización y evaluación de la salinidad. *Instituto de Suelo. Ministerio de La Agricultura*, (537), 1–9. Retrieved from



<http://www.fao.org/docs/eims/upload/cuba/5420/salinidad.pdf>

- Palomo Rodríguez, M., Rivera González, M., Martínez Rodríguez, G., Faz Contreras, R., & Espinoza Arellano, J. D. J. (2013). Uso de aguas salinas y frecuencias de aplicación asociado a la salinización de suelos y producción de algodónero. *Agrofaz*, 13(3), 119–124.
- Peréz, C., & Muñoz, Á. L. (2006). *Teledetección: nociones y aplicaciones. Journal of Chemical Information and Modeling* (Vol. 53).
<https://doi.org/10.1017/CBO9781107415324.004>
- SEMARNAT, S. (2010). *Salinidad del suelo*.
- Shahabi, M., Jafarzadeh, A. A., Neyshabouri, M. R., Ghorbani, M. A., & Kamran, K. V. (2017). Spatial modeling of soil salinity using multiple linear regression , ordinary kriging and artificial neural network methods. *Archives of Agronomy and Soil Science*, 63(2), 151–160. <https://doi.org/10.1080/03650340.2016.1193162>
- Yang, X., & Yu, Y. (2017). Estimating Soil Salinity Under Various Moisture Conditions: An Experimental Study. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5), 2525–2533. <https://doi.org/10.1109/TGRS.2016.2646420>
- Yue, J., Feng, H., Yang, G., & Li, Z. (2018). A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy. <https://doi.org/10.3390/rs10010066>



ANEXOS

Anexo 1: Población de Técnicas de estimación y la muestra escogida

Tabla 25

Población de técnicas de estimación

TÉCNICAS DE ESTIMACIÓN	PRECISIÓN	
	R ²	Error relativo
1. PARTIAL LEAST SQUARE REGRESSION (PLSR)	89.00%	
2. MULTIPLE LINEAR REGRESSION (MLR)	87.00%	
	52.00%	
	36.00%	
	96.30%	
	65.00%	
3. RANDOM FOREST REGRESSION (RF)	79.00%	
	98.30%	
	85.00%	
4. SUPPORT VECTOR MACHINE (SVM)	83.00%	
	61.70%	37.00%
5. BOOSTED BINARY REGRESSION (BBRT)	82.00%	
6. ARTIFICIAL NEURONAL NETWORK (ANN)	81.00%	
	93.00%	
	69.00%	
	99.70%	
	41.90%	
		49.00%
7. DECISION TREE REGRESSION (DT)	80.00%	
8. PRINCIPAL COMPONENT REGRESSION (PCR)	79.00%	

Fuente: Elaboración propia

Recopilado de:

- A. “ A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy “ (Yue, Feng, Yang, & Li, 2018)
- B. “Mapping and Modelling Spatial Variation in Soil Salinity in the Al Hassa Oasis Based on Remote Sensing Indicators and Regression Techniques” (Allbed, Kumar, and Sinha 2014)



C. “Dynamic prediction of soil salinization in an irrigation district based on the support vector machine” (Guan et al. 2013)
D. “Digital Mapping of Topsoil Salinity Using Remote Sensing Indices in Agh-Ghala Plain, Iran” (Mousavi et al. 2017).
E. “Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods”(Shahabi et al., 2017)
F. "Estimation of biomass in wheat using random forest regression algorithm and remote sensing data" (Wang et al. 2016)
G. "The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data" (Kayri, Kayri, and Gencoglu 2017)

El orden de las técnicas de estimación se extrajo de la fuente **A**, en dicha fuente se compara todas las técnicas de la población.

Anexo 2: Pasos para recolectar las imágenes multiespectrales

- 1. Para descargar las imágenes se accede al siguiente sitio web:

<https://earthexplorer.usgs.gov/>

- 2. En el primer menú se insertó el lugar **Chiclayo** a buscar las imágenes

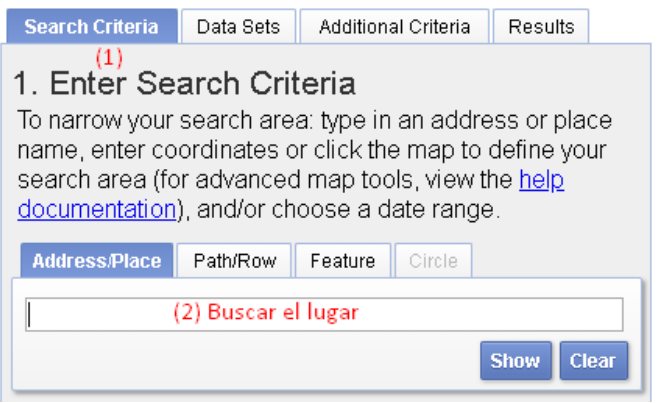


Figura 18: Búsqueda De Imágenes Multiespectrales Por Lugar.
Fuente: Elaboración Propia.



- Luego se buscó las imágenes desde **01 de enero de 2018 hasta 30 de setiembre de 2018**.

The screenshot shows the 'Search Criteria' tab of a search interface. It includes sections for 'Address/Place', 'Coordinates', and 'Date Range'. The 'Date Range' section is active, showing a search from '01/01/2018' to '09/30/2018' with '(3) Fechas' (3 dates) indicated. There are also buttons for 'Data Sets >', 'Additional Criteria >', and 'Results >'.

Figura 19: Búsqueda De Imágenes Multiespectrales Por Fecha.
Fuente: Elaboración Propia.

- En el segundo menú (DataSets) se seleccionó los orígenes de datos **Landsat 8 OLI/TIRS**.

The screenshot shows the 'Data Sets' tab of the search interface. It displays a tree view of data sets under the 'Landsat' category. The 'Landsat 8 OLI/TIRS C1 Level-1' option is selected and highlighted with a red box. Other options include Landsat 7 ETM+ C1 Level-1, Landsat 4-5 TM C1 Level-1, and Landsat 1-5 MSS C1 Level-1.

Figura 20: Búsqueda De Imágenes Multiespectrales Por Satélite.
Fuente: Elaboración Propia.



- En el tercer menú (Additional Criteria), se seleccionó **less than 30%** en **Scene Cloud Cover** para obtener las imágenes que tengan cobertura de nubes menor al 30%.

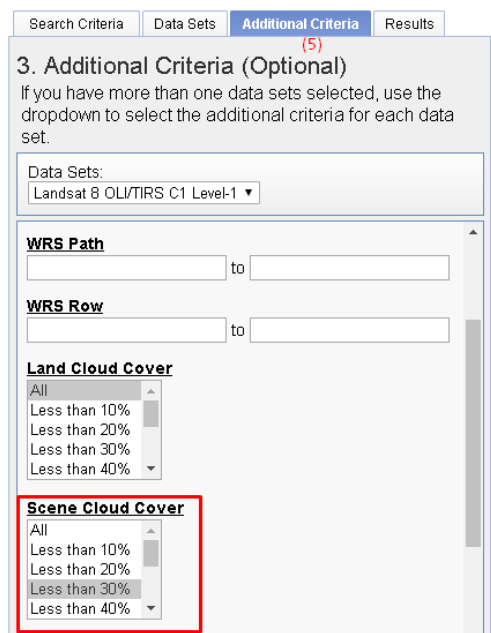


Figura 21: Búsqueda De Imágenes Multiespectrales Por Filtro De Porcentaje De Nubosidad.
Fuente: Elaboración Propia.

- En el cuarto menú (Results), se descargó las imágenes.

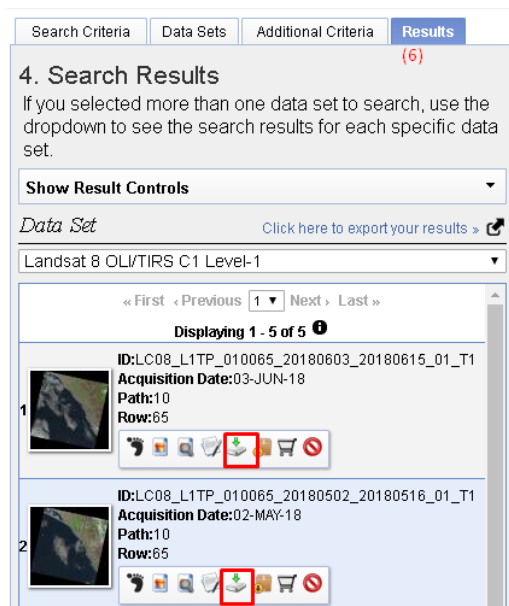


Figura 22: Vista De Adquisición De Imágenes Multiespectrales.
Fuente: Elaboración Propia.



- Por último se seleccionó el paquete **Level 1- GeoTIFF Data Product** para su descarga.

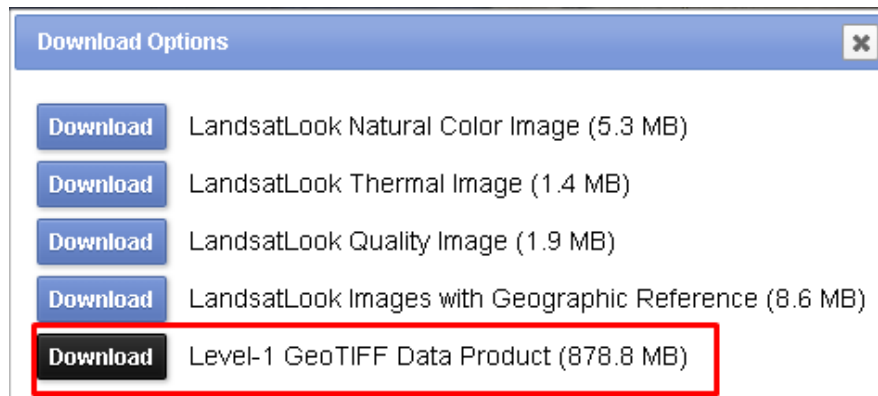


Figura 23: Adquisición De Paquete De Imágenes Multiespectrales.
Fuente: Elaboración Propia.

Anexo 3: Pasos para convertir los valores digitales a reflectancia manualmente.

- Ubicar y leer el archivo de metadata, es un archivo que tiene el sufijo **_MTL** al final y con extensión **.TXT**.

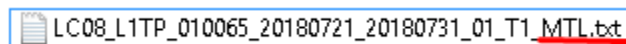


Figura 24: Archivo De Metadata De Un Paquete De Imágenes Multiespectrales
Fuente: Elaboración Propia.



2. Buscar las variables de REFLECTANCE_MULT_BAND_* y REFLECTANCE_ADD_BAND_*.

```
REFLECTANCE_MAXIMUM_BAND_1 = 1.210700
REFLECTANCE_MINIMUM_BAND_1 = -0.099980
REFLECTANCE_MAXIMUM_BAND_2 = 1.210700
REFLECTANCE_MINIMUM_BAND_2 = -0.099980
REFLECTANCE_MAXIMUM_BAND_3 = 1.210700
REFLECTANCE_MINIMUM_BAND_3 = -0.099980
REFLECTANCE_MAXIMUM_BAND_4 = 1.210700
REFLECTANCE_MINIMUM_BAND_4 = -0.099980
REFLECTANCE_MAXIMUM_BAND_5 = 1.210700
REFLECTANCE_MINIMUM_BAND_5 = -0.099980
REFLECTANCE_MAXIMUM_BAND_6 = 1.210700
REFLECTANCE_MINIMUM_BAND_6 = -0.099980
REFLECTANCE_MAXIMUM_BAND_7 = 1.210700
REFLECTANCE_MINIMUM_BAND_7 = -0.099980
REFLECTANCE_MAXIMUM_BAND_8 = 1.210700
REFLECTANCE_MINIMUM_BAND_8 = -0.099980
REFLECTANCE_MAXIMUM_BAND_9 = 1.210700
REFLECTANCE_MINIMUM_BAND_9 = -0.099980

SUN_ELEVATION = 49.77414932
```

Figura 25: Variables De Metadata A Utilizar Para Calcular La Reflectancia
Fuente: Elaboración Propia.

3. Crear tablas en base a las variables anteriores de la banda 1-7.

$$p_{\lambda'} = M_p Q_{cal} + A_p$$

$$p_{\lambda} = \frac{p_{\lambda'}}{\sin(\theta_{SE})}$$



Anexo 4: Programa para convertir los valores digitales a reflectancia automáticamente.

- Métodos para leer el archivo de metadata

```
def buscarMetadata(self,cadena):
    nlist = []
    for el in self.lista:
        c = self.buscarPalabra(cadena,el)
        if c is not None:
            nlist.append(self.buscarPalabra(cadena,el))
    return nlist

def buscarPalabra(self,c,p):
    if filter(lambda x: c in x, p):
        return p

def extraerBandas(self):
    busq = self.buscarMetadata("REFLECTANCE_ADD_")
    for el in busq:
        temp = el[0].replace("REFLECTANCE_ADD_", "")
        self.bandas.append(temp)

def leerMetadata(self, ruta):
    metadata = open(ruta,"r")
    lines = metadata.readlines()
    for line in lines:
        line = line.replace(' ','')
        line = line.replace('\n','')
        line = line.replace("'",'')
        tline = line.split("=")
        self.lista.append(tline)
    self.extraerBandas()

    metadata.close
```

Figura 26: Código Fuente En Python Para Leer El Archivo De Metadata
Fuente: Elaboración Propia



- Métodos para leer los pixeles de la imagen en una matriz numérica, y calcular la reflectancia.

```
def leerImagen(self):
    dataset = gdal.Open(self.ruta)
    self.metadata = dataset.GetGeoTransform()
    band = dataset.GetRasterBand(1)
    transform = dataset.GetGeoTransform()
    self.xOrigin = transform[0]
    self.yOrigin = transform[3]
    self.pixelWidth = transform[1]
    self.pixelHeight = -transform[5]
    matriz = band.ReadAsArray(0, 0, dataset.RasterXSize, dataset.RasterYSize).astype(numpy.float)
    self.calcularReflectancia(matriz)
```

Figura 27: Código Fuente En Python De La Función Para Leer Una Imagen

Fuente: Elaboración Propia

```
def calcularReflectancia(self,matriz):
    sun_correction = math.sin(math.radians(self.banda.SUN_ELEVATION))
    #sun_correction = numpy.sin(self.banda.SUN_ELEVATION)
    self.reflectancia = ((matriz * self.banda.REFLECTANCE_MULT_BAND)+ self.banda.REFLECTANCE_ADD_BAND)/sun_correction
```

Figura 28: Código Fuente En Python Para Calcular La Reflectancia

Fuente: Elaboración Propia

- Extracción de valores en metadata

```
from metadata import Metadata
from imagen import Imagenes,Imagen,Banda
from osgeo import gdal
import matplotlib.pyplot as plt
import sys
import math
import numpy as np

if __name__ == '__main__':
    directorio = "C:\\Users\\User\\Desktop\\IMAGENES MULTIESPECTRALES\\20180126_20180207"
    ruta = directorio+"\\LC08_L1TP_010065_20180126_20180207_01_T1_MTL.txt"
    metadata = Metadata()
    metadata.leerMetadata(ruta)
    imagenes = Imagenes()
```

Figura 29: Código Fuente En Python Para Calcular La Reflectancia

Fuente: Elaboración Propia



- Cálculo de la reflectancia de las imágenes espectrales.

```

for ban in metadata.bandas[:7]:
    ruta = metadata.buscarMetadato("FILE_NAME_" + ban)
    if ruta is not None:
        banda = Banda()
        banda.nombre = ban
        banda.REFLECTANCE_MULT_BAND = float(metadata.buscarMetadato("REFLECTANCE_MULT_" + ban)[0][1])
        banda.REFLECTANCE_ADD_BAND = float(metadata.buscarMetadato("REFLECTANCE_ADD_" + ban)[0][1])
        banda.SUN_ELEVATION = float(metadata.buscarMetadato("SUN_ELEVATION")[0][1])
        img = Imagen()
        img.banda = banda
        img.ruta = directorio + '\\' + ruta[0][1]
        imagenes.lista.append(img)

for img in imagenes.lista:
    print "Leyendo imagen" + img.banda.nombre
    img.leerImagen()
    
```

Figura 30: Código Fuente En Python De La Ejecución De Las Funciones De Leer De Imagen Y Leer La Metadata

Fuente: Elaboración Propia

Anexo 5: Pasos para calcular los indicadores manualmente.

1. Identificar el indicador a utilizar, y las imágenes que se utilizaran.

$$NDVI = \frac{p_{NIR} - p_R}{p_{NIR} + p_R}$$

2. Las bandas a utilizar para este indicador son las p_{NIR} (Imagen de la reflectancia en la banda infrarrojo cercano o Banda 6) y p_R (Imagen de la reflectancia en la banda roja o Banda 5).

Se obtiene como referencia un pixel de la imagen, y el valor de dicho pixel representa la reflectancia en dicha zona.

El valor de reflectancia para la **Banda 6** (p_{NIR}) es **0.1752**

El valor de reflectancia para la **Banda 5** (p_R) es **0.1618**

3. Reemplazar las reflectancias obtenidas en la ecuación del indicador.

$$NDVI = \frac{p_{NIR} - p_R}{p_{NIR} + p_R}$$

$$NDVI = \frac{0.1752 - 0.1618}{0.1752 + 0.1618}$$

$$NDVI = \frac{0.0134}{0.337}$$

$$NDVI = 0.04$$



Anexo 6: Programa para calcular los indicadores de una imagen automáticamente.

- Referenciación de las imágenes y el identificador de reflectancia.

```
pblue = imagenes.buscarImagen("BAND_2")
pgreen = imagenes.buscarImagen("BAND_3")
pred = imagenes.buscarImagen("BAND_4")
pnir = imagenes.buscarImagen("BAND_5")
pswir = imagenes.buscarImagen("BAND_6")
```

Figura 31: Código Fuente En Python De Lectura De Una Imagen Espectral Procesada
Fuente: Elaboración Propia

- Implementación de los indicadores.

```
pndvi = (pnir-pred)/( pnir + pred)
```

Figura 32: Código fuente en Python implementación del indicador NDVI
Fuente: Elaboración propia

- Copia de la metadata (georreferenciación) de la imagen.

```
output_image = Imagen()
imt = imagenes.lista[0]
output_image.xOrigin = imt.xOrigin
output_image.yOrigin = imt.yOrigin
output_image.pixelWidth = imt.pixelWidth
output_image.pixelHeight = imt.pixelHeight
output_image.metadata = imt.metadata
```

Figura 33: Código Fuente En Python De Lectura De Los Puntos De Origen Y El Tamaño De Pixel De Una Imagen Espectral
Fuente: Elaboración Propia

- Cálculo del indicador.

```

directorio = directorio + "\\INDICADORES\\"
print "Exportando NDVI"
output_image.guardarImagen(directorio + '\\NDVI.tif',pndvi)
    
```

Figura 34: Código Fuente En Python De La Exportación De Un Indicador Calculado.
Fuente: Elaboración Propia.

Anexo 7: Pasos para ubicar una coordenada en una imagen con información geoespacial, mediante una interfaz gráfica.

1. La ecuación para obtener la ubicación (fila y columna) de un pixel en una imagen (m filas y n columnas), mediante la coordenada de este es la siguiente.

$$fila = \frac{A - X_0}{m}$$

$$columna = \frac{Y_0 - B}{n}$$

Donde:

A: es la coordenada horizontal a buscar

X_0 : La coordenada inicial horizontal

m: Cantidad de pixeles horizontales

B: es la coordenada vertical a buscar

Y_0 : La coordenada inicial horizontal

n: Cantidad de pixeles verticales

2. Identificar la coordenada a ubicar y la imagen donde se ubicará, para extraer el valor de dicho pixel.

La ecuación para convertir una coordenada UTM 17S A UTM 17N es la siguiente:

$$(X, Y)$$

$$X_N = X_S$$

$$Y_N = Y_S - 1000000$$



Donde:

X: Es la coordenada horizontal

Y: Es la coordenada vertical

X_N, Y_N : Coordenadas UTM 17N

X_S, Y_S : Coordenadas UTM 17S

Tabla 26

Información de coordenada a buscar y el indicador donde se buscará

COORDENADA UTM 17S	IMAGEN
617226 , 9265494	NDVI

Fuente: Elaboración Propia

Conversión de la coordenada

$$617226, 9265494 \text{ (UTM 17S)} = 617226, -734506 \text{ (UTM 17N)}$$

Tabla 27

Ubicación y valor de un pixel en una imagen espectral

A	B	X_0	Y_0	m	n
617226	-734506	495000	-683400	30	30

Fuente: Elaboración Propia

3. Reemplazar la coordenada en la fórmula.

$$fila = \frac{A - X_0}{m}$$

$$fila = \frac{617226 - 495000}{7611}$$

$$fila = 4074$$

$$columna = \frac{Y_0 - B}{n}$$

$$columna = \frac{-683400 - -734506}{30}$$

$$columna = 1703$$



Anexo 8: Programa para ubicar coordenadas en una imagen con información geoespacial.

- Función que lee la información de una imagen, valor de pixel en valores numéricos, y la metadata (coordenadas asociadas a la imagen),

```
def leerIndice(self):
    dataset = gdal.Open(self.ruta)
    band = dataset.GetRasterBand(1)
    transform = dataset.GetGeoTransform()
    self.xOrigin = transform[0]
    self.yOrigin = transform[3]
    self.pixelWidth = transform[1]
    self.pixelHeight = -transform[5]
    self.reflectancia = band.ReadAsArray(0, 0, dataset.RasterXSize, dataset.RasterYSize).astype(numpy.float)
```

Figura 35: Código Fuente En Python De Lectura De Un Indicador

Fuente: Elaboración Propia

- Función que convierte la coordenada de formato UTM 17S (variable epsg:32717) a UTM 17N (variable epsg:32617), luego busca la posición en la imagen y extrae el valor del pixel.

```
def buscarPixel(self, coordenadas):
    valores = []

    for point in coordenadas:
        x, y = proj.transform(proj.Proj(init='epsg:32717'), proj.Proj(init='epsg:32617'), point[0], point[1])
        row = int((x - self.xOrigin) / self.pixelWidth)
        col = int((self.yOrigin - y) / self.pixelHeight)
        reflec = self.reflectancia[col][row]
        valores.append(reflec)
        print reflec
    return valores
```

Figura 36: Código Fuente En Python De Lectura De Coordenada, Conversión Y Búsqueda De Coordenada.

Fuente: Elaboración Propia



- Ejecución de las funciones anteriores.

```
array = ('RED', 'GREEN', 'BLUE', 'NDVI', 'DVI', 'BI', 'ASTERSI', 'SI', 'SI1', 'SI2', 'SI3', 'SI11')
imagenes = Imagenes()

df = pd.DataFrame()

cont = 0
for elmnt in array:
    path = directorio + '\\' + elmnt + ".tif"
    img = Imagen()
    print path
    img.ruta = path
    img.leerIndice()
    points_list = [
        (617218,9265607),
        (617211,9265623),
        (617183,9265653),
        (617156,9265671),
        (617153,9265644),
        (617140,9265616),
        (617208,9265589),
        (617196,9265572),
        (617180,9265587),
        (617171,9265583)
    ]
```

Figura 37: Código Fuente En Python De Ejecución De Búsqueda De Coordenadas Y Extracción De Indicadores.

Fuente: Elaboración Propia

- Exportar la información obtenida a un archivo csv

```
df.insert( loc=cont, column='EC' , value=[0.9,1.22,0.84,1.5,1.6,1,1.4,1.78,1.6,1.46])
print df
df.to_csv( directorio + '\\' + "caracteristicasMAYASOFT.csv", sep=',', encoding='utf-8')
```

Figura 38: Código Fuente En Python De Ejecución De Búsqueda De Coordenadas Y Extracción De Indicadores.

Fuente: Elaboración Propia



- Ejecución de la Técnica Regresión Lineal Múltiple

```
from sklearn import linear_model

if __name__ == '__main__':

    dataset = pd.read_csv('C:\\Users\\User\\Desktop\\IMAGENES MULTIESPECTRALES\\20180126_20180207\\INDICADORES\\caracteristicasaMAYASOFT.csv',",")

    indicadores = ['RED','GREEN','BLUE','NDVI','DVI',
                  'BI','ASTERSI','SI','SI1','SI2','SI3','SI11']

    X = dataset[indicadores]
    y = dataset['EC']
    regr = linear_model.LinearRegression()
    regr.fit(X, y)
    y_pred = regr.predict(X)
    print 'R Squared (R2):', metrics.r2_score(y, y_pred)
    print 'Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y, y_pred))
    df = pd.DataFrame({'CE Real': y, 'CE Estimada': y_pred})
    print df
```

Figura 39: Código Fuente En Python De Ejecución De La Técnica Regresión Lineal Múltiple
Fuente: Elaboración Propia

- Ejecución de la Técnica Random Forest

```
from sklearn.ensemble import RandomForestRegressor

if __name__ == '__main__':

    dataset = pd.read_csv('C:\\Users\\User\\Desktop\\IMAGENES MULTIESPECTRALES\\20180126_20180207\\INDICADORES\\caracteristicasaMAYASOFT.csv',",")

    indicadores = ['RED','GREEN','BLUE','NDVI','DVI',
                  'BI','ASTERSI','SI','SI1','SI2','SI3','SI11']

    X = dataset[indicadores]
    y = dataset['EC']
    regr = RandomForestRegressor(max_depth=7, n_estimators=500)
    regr.fit(X, y)
    y_pred = regr.predict(X)
    print 'R Squared (R2):', metrics.r2_score(y, y_pred)
    print 'Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y, y_pred))
    df = pd.DataFrame({'CE Real': y, 'CE Estimada': y_pred})
    print df
```

Figura 40: Código Fuente En Python De Ejecución De La Técnica Random Forest
Fuente: Elaboración Propia



Anexo 9: Análisis de muestras de suelos con humedad, mediante conductividad eléctrica.



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE AGRONOMIA
Oficina de Administración



ANALISIS QUIMICO DE SUELOS

SOLICITANTE : CHRISTIAN OMAR AMAYA DIAZ

TESIS : COMPARACION DE TECNICAS DE ESTIMACION DEL GRADO DE SALINIDAD EN SUELOS CON CONTENIDIO DE HUMEDAD MEDIANTE EL PROCESAMIENTO DE IMAGENES MULTIESPECTRALES

PROCEDENCIA : DISTRITO MORROPE, PROVINCIA LAMBAYEQUE, REGION LAMBAYEQUE

ASUNTO : ANÁLISIS QUÍMICO DE DIEZ (10) MUESTRAS TRAÍDOS POR EL INTERESADO AL LABORATORIO DE SUELOS.

Muestra N°	C.E _s dS/m
A1	0.90
A2	1.22
A3	0.84
A4	1.50
A5	1.60
A6	1.00
A7	1.40
A8	1.78
A9	1.60
A10	1.46

Lambayeque, 19 de octubre del 2018



JEFATURA DE LABORATORIO DE SUELOS

Figura 41: Análisis De Muestras De Suelos Con Humedad, Mediante Conductividad Eléctrica.
Fuente: Elaboración Propia

